Optimal Prediction, Alarm, and Control in Buildings Using Thermal Sensation Complaints

by

Rodney Alexander Martin

B.S. (Carnegie-Mellon University 1992)
M.S. (University of California at Berkeley 2000)

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering - Mechanical Engineering

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor David M. Auslander, Chair
Professor Alice Agogino
Professor Edward Arens

Fall 2004

The dissertation of Rodney Alexander Martin is approved:

Chair Date

Date

Date

University of California, Berkeley

Fall 2004

Optimal Prediction, Alarm, and Control

in Buildings Using Thermal Sensation Complaints

by

Rodney Alexander Martin

Abstract

Optimal Prediction, Alarm, and Control

in Buildings Using Thermal Sensation Complaints

by

Rodney Alexander Martin

Doctor of Philosophy in Engineering - Mechanical Engineering

University of California, Berkeley

Professor David M. Auslander, Chair

The management and automation of commercial building HVAC (Heating, Ventilation and Air-Conditioning) systems is illustrative of a business that can benefit from the insightful use of all available information sources. Modern HVAC systems using direct digital control can potentially provide useful performance data. Building occupant feedback complaint statistics are an untapped resource, and can be cultivated from modern maintenance management databases. This thesis will address the integration and application of these fundamental sources of information, using some modern and novel techniques. The cost and scalability of these techniques can be positively influenced by recent technological advances in computing power, sensors, and databases. However, application of the theory that drives some of the newer techniques is still in its infancy, and requires tremendous computing power. As such, an important theme in the thesis is to address computational efficiency and practical usefulness of the techniques, via some clever approximations. One desired outcome is to make these new techniques more accessible to users of existing building control technology. It is hoped that these users and facility managers as well can benefit from the algorithmic and design-oriented paradigm

in which the results are presented.

Professor David M. Auslander
Dissertation Committee Chair

Dedicated to my Parents: June Carol and Bartholomew Martin, Jr.

Contents

List of Figures

List of Tables

List of Symbols

Standard Math and Set Symbols

Symbol	Description
$\lvert \cdot \rvert$	Absolute Value or Cardinality (number of elements in a set)
Δ	Difference Operator
$\approx$	Approximately
$\gtrapprox$	Approximately greater than
$\pm$	Plus or minus
$\Longleftrightarrow$ or $\Leftrightarrow$	Formally: "if and only if (iff)," Loosely: equivalent to
$\Rightarrow$ or $\Longrightarrow$	Formally: "implies," Loosely: "Therefore"
$\succ$	Positive definite
$\succeq$	Positive semi-definite
$\prec$	Negative definite
$\preceq$	Negative semi-definite
:	such that
$\in$	is a member of the set
$\forall$	For all
$\exists$	There exists
$\triangleq$	is defined as
$\equiv$	is identical to
$\cap$ or $\bigcap$	Intersection of sets (as an unary or binary operator)
$\cup$ or $\bigcup$	Union of sets (as an unary or binary operator)
$\subset$	contained in the set
$\subseteq$	is a subset of
$\nsubseteq$	is *not* a subset of
$\{A\}$	The set A
$\{A : B\}$	The set A given B
$\emptyset$	Null Set

Standard Math and Set Symbols, Continued

Symbol	Description
v	Generic real scalar
$\mathbf{v}$ or $\mathbf{V}$ ($\mathbb{V}$)	Generic real vector or matrix, respectively
$\mathbf{v}^T$ or $\mathbf{V}^T$	Transpose of a generic vector or matrix, respectively
$\hat{v}$ or $\hat{\mathbf{v}}$ or $\hat{\mathbf{V}}$	Relative frequencies or (un)conditional estimates of a generic scalar, vector, or matrix, respectively
$\mathcal{O}(\cdot)$	Order of magnitude of computational burden/complexity
$rank(\mathbf{A})$	Rank of matrix $\mathbf{A}$
Ω_A	The region or area, A
$\mathbb{R}$	The set of real scalar numbers
$\mathbb{R}^n$	The set of real column vectors having n real numbers
$\neg A$ or A'	Complement of set A
J' or γ'	$'$ Can also be used as "prime," in the proper context
$\setminus$	Set subtraction (take away)
$\lfloor \cdot \rceil$	Value rounded to the nearest integer
$\|v\|_1 = \sum_{i=1}^{n} \|v_i\|$	1-norm
$v^{opt} = \arg\min_v J(v)$	The optimal minimizing v for the cost function $J(v)$
$v^{opt} = \arg\max_v J(v)$	The optimal maximizing v for the cost function $J(v)$

Probabilistic Symbols

Symbol	Description
$\sim$	is distributed as
$P(A)$	Probability of A
$P(A\|B)$	Probability of A, given B
$E[A]$	Expectation or expected value of A
$E[A\|B]$	Conditional expectation or expected value of A, given B
σ	Standard Deviation
μ	Mean
ν	Frequency
σ^2 or $Var(\cdot)$	Variance
$cov(\cdot,\cdot)$	Covariance
$cor(a,b) = \frac{cov(a,b)}{\sigma_a \sigma_b}$	Correlation
$\mathbf{\Sigma}$	Covariance Matrix
$\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$	Multivariate Normal distribution with mean $\boldsymbol{\mu}$ and Covariance Matrix $\mathbf{\Sigma}$
$\phi(v) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{v^2}{2}}$	Standard Normal Probability Density Function
$\Phi(v) = \int_{-\infty}^{v} \phi(z)dz$	Cumulative Normal Standard Distribution Function
$\Phi^{-1}(v)$	Inverse Cumulative Normal Standard Distribution Function
$Var(A\|B)$	Variance of A conditioned on B
$cov(A,B\|C)$	Covariance of A,B conditioned on C

Application-Specific Symbols

Symbol	Description
^oC	Degrees Celsius
^oF	Degrees Fahrenheit
T	Temperature
T_B	Building temperature
T_H	Hot complaint process temperature
T_L	Cold complaint process temperature
T_h	Hot complaint event temperature
T_l	Cold complaint event temperature
$v(t)$	Generic continuous-time process
$\dot{v}(t)$	Derivative of generic continuous-time process
v_k	Generic discrete-time process
ω	Natural frequency
ρ	Autocorrelation
τ or τ_s or T_s	Sampling interval
λ	Forgetting factor
$\boldsymbol{\lambda}$	Lagrange multiplier
$\mathbf{V}_{ss}$	Generic steady-state value of $\mathbf{V}$
$\mathbf{V}^L$	Generic Lyapunov value of $\mathbf{V}$
$\mathbf{V}^R$	Generic Riccati value of $\mathbf{V}$

List of Abbreviations

Acronym	Description
AC	Air–Conditioning
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANSI	American National Standards Institute
ASHRAE	American Society of Heating, Refrigeration, and Air–Conditioning Engineers
CDF	Cumulative Distribution Function
CEC	California Energy Commission
CFR	Code of Federal Regulations
DARE	Discrete Algebraic Riccati Equation
DDC	Direct Digital Control
EM	Expectation-Maximization
EWMA	Exponentially-Weighted Moving Average
FOH	First-Order Hold (or linear interpolation)
G-S	Gaussian-Stationary
HVAC	Heating, Ventilation and Air-Conditioning
IESNA	Illuminating Engineering Society of North America
ISO	International Organization for Standardization
kW-h	kilowatt-hour
LQ	Linear-Quadratic
PDF	Probability Density Function
PI	Proportional-Integral
PID	Proportional-Integral-Derivative
PMV	Predicted Mean Vote
PPD	Predicted Percent Dissatisfied
TMY	Typical Meteorological Year
Type I error	Probability of False Alarm
Type II error	Probability of Missed Detection
VAV	Variable Air Volume
ZOH	Zero-Order Hold

Acknowledgments

Over the past 7 years here at Berkeley, I have not simply obtained two graduate degrees, but I've also learned and grown immeasurably as a student, researcher, and as a person in ways that I never imagined possible at the commencement of my studies. I've found that there are soaring highs and extreme lows in the course of completing a comprehensive graduate education. However, I celebrated each victory, no matter how small, and I learned from each setback as well, no matter how minor. Because no graduate education is complete without the help of others, I would like to thank so many of the diverse group of people that have helped me throughout my long journey. My experience at Berkeley has definitely had an international flair. I've made friends with people of different ethnicities and countries often who speak several different languages, who come from different cultures and backgrounds, and I would have it no other way.

First I would like to thank Dr. Carla Trujillo, formerly Director of the Graduate Academic Diversity Program from the College of Engineering, who is now the Director of the Graduate Opportunity Program for the entire University. Without her unwavering encouragement and financial support, my graduate studies would have most certainly been harder to navigate, and prevented me from pursuing my true research and academic goals. I would next like to thank my dissertation chair, and co-advisor, Professor David Auslander and Dr. Clifford Federspiel respectively, who have taught me the most valuable lesson of all about engineering research: to respect the research question by selecting appropriate tools to answer it, rather than letting the tools drive the research, now matter how clever they are. I also thank them both for their patience, support, objectivity, and constructive criticism throughout the course of my search for the research question and its answer. Cliff, thanks for planting the seed of how to approach the details of research in the building controls domain and for introducing me to the uniquely rich and interesting field of level-crossing theory. I am indebted to you for helping me find summer support on so many occasions, even once at LBNL (Lawrence Berkeley National Laboratory) with Dr. Phil Haves and Dr. Peng Xu. Thanks also for helping me to develop a research posture based on our common interests, and allowing me to branch

off from your previous work.

I would also like to thank so many of the folks that I've worked with at the Center for Environmental Design Research over the years for their enthusiasm, encouragement, and support. This includes all of the students, faculty, and staff members, both researchers and support staff that have been so friendly and helpful during my tenure here. A few of the current and former names over the years are: Charlie Huizenga, Tom Webster, Fred Bauman, Zhang Hui, Prof. Gail Brager, Danni Wang, Erik Ring, Huilin Li, Hannah Yan, Luis Villafana, Jane Lin, Gwelen Paliaga, Therese Pfeffer, Keke Chen, Robert Marcial, Kevin Powell, David Lehrer, Karen Roy, Nora Watanabe, and so many others who I may have failed to mention. I also thank the Director of the Center, Prof. Edward Arens, for agreeing to serve as a member of my dissertation committee. Thanks also to Prof. Agogino from the Department of Mechanical Engineering, for not only being a member of my dissertation committee, but for serving on my qualifying exam committee as well, and fostering a collaborative research environment within the department.

There are so many students that I've befriended over the years, some who've come and gone, and others that are still here, that I would be remiss in not mentioning them. I would also like to thank Prof. Andy Packard for serving as a member of my Masters and qualifying exam committees, and giving me half-time support to work on some interesting research issues at the BCCI (Berkeley Center for Control and Identification) lab in 5105 Etcheverry Hall during the summer of '99. Thanks also for allowing me to stay there for such a long time afterwards without actually being involved in the research of the lab. Being able to spend time with the students and professors of the lab certainly enriched and diversified my experience here not only academically and professionally, but personally. To name a few of the current and former members during my stay: Prof. Kameshwar Poolla, Khalid Al-Ali (K.J.), Eric Wemhoff, Mareike Claassen, Dave Faris, Mason Freed, Darin Fisher, Alpay Kaya, Lun Li, Kunpeng Sun, Zachary Jarvis-Wloszek, Michiel Krüger, George Panayotopoulos, Lance Hazer, Douglas Philbrick, Andrew Flores, Matt Onsum, Peter Sedivec, Weehong Tan, and others who I may have failed to mention.

There are also the former and current students of Professor Auslander's lab in 1176A Etcheverry Hall that I have either worked with closely or were labmates with during the final years of my tenure here. To name them: John and Pam Ridgely, Chris Pawlowski, Barton Clark, Fayez Khan, Sasha Lanning, Craig Lin, Tim Boucher, Shan Tang, Xue Chen, Jaehwi Jang, Phil Cutino, Jørgen Johnsen, Gregory Burton, Mauricio Delapeña, Steve Sulack, Dean Tsai, Josh Cemenska, Andy Jennings, and others that I may have failed to mention. Although not currently belonging to any of the labs I've mentioned, I would like to thank other current and former students, providing support in classes and preparation for preliminary and qualifying exams, etc.: Richard Liu, Marcus McElhaney, Duane Edmonds, Pedro Arroyo, Gene Shin, Dennice Gayme, Shige Kitani, Chris Foley, Matt Giere, Deborah Jue, Jessica Granderson, Michael Uchanski, Gabriel Gomes, Pete Seiler, Ahmed El-Shaer and others who I may have failed to mention.

Special thanks to Elaine Shaw, Priya Sreedharan, and Laura Muñoz for those commiserations, coffeehouse and hallway chats that allowed me to vent, from earlier days to those stressful final hours of research and writing. Thanks also to Phil Buonadonna for providing me a familiar face from the past here at Berkeley, and for helping me out during my transportation tragedy. Thanks to those in the Navy Reserves who have supported my cause in pursuit of this degree, especially Ralph Wilson and Mary Chen, as well as Suzanne Fiori and April Dixon. I would especially like to thank Jonghak Kim of Professor Auslander's lab for his friendship and support during our GSI work for ME230, and late night discussions on everything from exercise and diet techniques to research philosophies. I could not possibly end this section without mentioning a wonderful young lady from Germany who brought so much joy and passion for learning into the lab. Ms. Anuscheh Nawaz was a member of the lab for what seems too short of a time, and I am so thankful for having met you, and for your love and support. I am happy that you are in my life during this transition to a new challenge, even at such a great distance.

Last but definitely not least, I would like to thank my family, without whose unswerving understanding, love, patience, support, I would never have even had a shot at shooting for the stars.

To my siblings, Russell and family, Renee, Raina, and Riva, thank you for your support and your mere presence: it has been such a blessing to watch all of you grow into the adult world. And to my parents, thank you so much for allowing me to vent every weekend and then some, and being understanding and just being there for me...always.

Chapter 1

Introduction

1.1 Preliminaries

Energy savings and thermal comfort are important to both facility managers and building occupants. As a result, they often are open to new and innovative ways to improve or even replace currently existing practical methods that might not be taking advantage of the most recent advances in technology. On one hand, in large commercial buildings modern DDC (Direct Digital Control) systems are becoming more ubiquitous, with the use of sophisticated hardware that networks HVAC components together and can be monitored remotely from a central location. The general trend in the design and commissioning of new commercial buildings includes these new types of systems. However, there still exists a large contingent of older buildings that still use pneumatic HVAC systems. Many facility managers are charged with the operation of not only a hybrid mix of older pneumatic building controls and modern DDC, but different building types and systems as well. There are also certain nuances that need to be handled properly, such as the geographical climate, weather, seasonal patterns' influence on management of HVAC system operation, perimeter vs. core zones, as well as building occupancy trends due to varied shifts and operating schedules. Due to the deregulation of energy markets both domestically and abroad, the cost of energy and pricing/rate

structures is another variable within the vast array of issues in a complicated operation that all building managers and their technicians must contend with.

There are several different business processes as well as physical operating systems that exist in facility management which are often mapped onto supervisory management information systems. Hence there is often a repository of stove-piped information sources that are not necessarily linked and used as inter-dependently as they should, and when they are it is more often on an ad-hoc/heuristic basis. As a result, the opportunity for widespread potential cost savings may be lost due to lack of knowledge or research concerning the intelligent use of these information-rich sources. The most recent technological advances in computing must also be taken advantage of to achieve the desired objectives of reduced energy usage and improved building occupant thermal comfort. To address this, the primary goal of the next two chapters is to do the following:

- Describe the thesis objectives, specifically aimed closing the gap between current practice methodology and current research efforts.
- Illustrate some of the current issues in commercial building operations, and how they are practically dealt with in industry today.
- Discuss some of the current research efforts attempting to improve and/or optimize the status quo of building operations as it pertains to commercial building HVAC systems.

1.2 Thesis Objective

It is important from a practical standpoint that the solutions to the problems cited are not solely dependent on theoretical first–principles based models, but enhanced with actual measured data from the plant and building occupants as well. Building occupants can be used as sensors in a "feedback loop" of sorts, in conjunction with measured data from the plant. An example of such a plant is shown in Fig. 1.1, enclosed within the dotted line labelled continuous–time "inner-loop." This represents a typical closed-loop control system, complete with disturbances due to weather and

internal loads, sensor lag at the thermostat measurement points, and a standard PI (Proportional-Integral) controller to regulate the zone temperature to a desired reference value, r.

Building occupant feedback is represented by the outer loop, which involves the development of both a complaint model and a thermostat setting control policy to achieve the desired objectives. At the center of Fig. 1.1 is the building plant itself, whose parameters are typically derived from a theoretical first–principles based model. However, as an alternative to this, a model that is completely based upon the aggregate statistics of measured building temperature data (via datalogger or DDC system) can be used, but it should be sufficient to characterize the fundamental system dynamics. They should allow for the same objectives to be achieved, nearly as optimally as a theoretical model would be able to. This allows for reduced complexity in the design and analysis stage, primarily when developing the complaint model and optimal alarm system, which will be discussed at length in subsequent chapters.

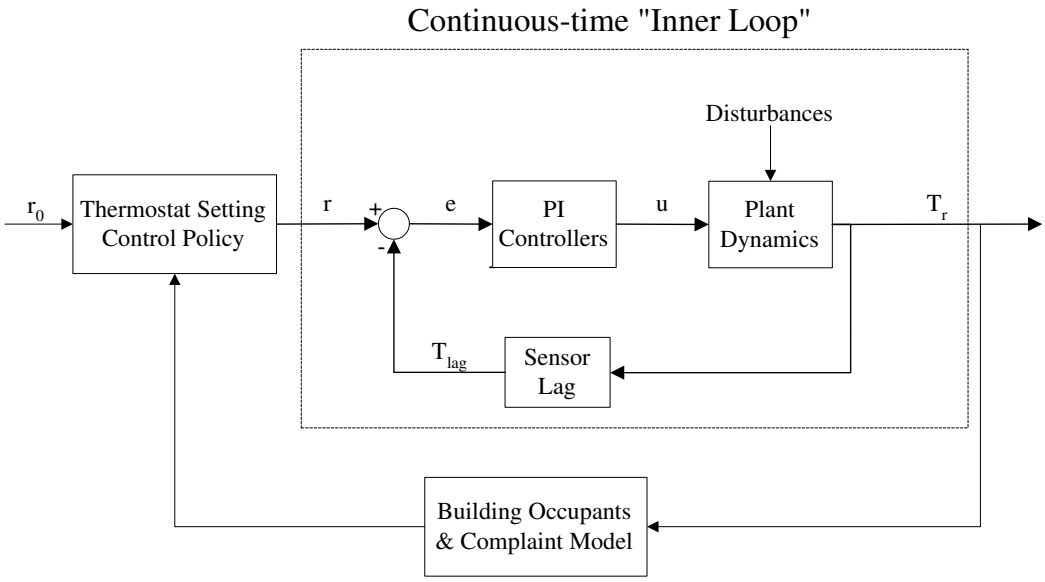

Figure 1.1: Block Diagram of Integrated System

It is necessary to obtain the statistics of measured complaint data, and the building plant in order to develop a baseline complaint model [19, 22]. These statistics, if updated in a real–time setting, would allow the predicted complaint model to achieve an adaptive on–line, real–time functionality. However, research indicates [46, 47, 21] that the complaint rates are often low, and the occupant feedback so sporadic that updating a complaint model adaptively may not have sufficient "persistence of excitation" required for model validity. In this case, where there is a lack of existing information to update the model, an off–line approach may be more suitable. However, there are other collateral ramifications of using occupant feedback, since building occupants cannot be modelled as perfect sensors. Occupants can learn to adjust their complaint behavior for a variety of psychological and sociological reasons. Hence we can only assume that measured data is based on normal behavior and any results of statistical information due to pathological behavior will have to be interpreted on an individual basis. Furthermore, we assume that the HVAC building systems are operating normally. This means that there are no faults in the system that might be the root cause of building occupant complaints. Typical faults would be leaky valves and components, HVAC equipment malfunction or failure, and miscalibrated sensors. Complaints are assumed to be solely due to building occupant disagreement with the standing facility operating policy, or the more probable case of poor control performance of any local zone controllers.

Given a model as described above with all of the stated assumptions in place, we must keep in mind the primary goal of any facility manager, which is to keep energy costs reasonable without sacrificing building occupant comfort. In fact, it may be useful for them to balance the existing energy budget against *anticipated* comfort levels. These comfort levels are clearly related to thermal sensation complaints, or events. In many engineering systems, the ability to predict, and give an accurate alarm prior to impending critical events is of great importance. These critical events may have varying degrees of severity, and in fact they may occur during normal system operation. An alarm may be given for any number of thermal sensation complaint events that occur over a specified time period. As such, an optimal alarm system can be designed to warn facility managers

of impending complaints that might occur within a specified time period, to aid them in making critical decisions about building operations. In this thesis, optimal alarm systems will be covered in detail in Chap. 3, and the results of simulated alarm systems will be presented as well, in Chap. 7.

Level-crossing formulae are used extensively in optimal alarm and control design in this thesis. New approximations that establish useful relationships between discrete-time optimal control and alarm theory and continuous time-level crossing formulae will be presented in Chap. 4. Such approximations may not have been considered in the past because there currently exist discrete-time level-crossing formula that provide exact results. However, in specific cases using these formulae are computationally burdensome and therefore not very practical. This is true especially for the implementation of real-time optimal control algorithms, such as one that might be used for a thermostat setting control policy shown in Fig. 1.1. Some very straightforward approximations can provide an alternative to the exact discrete-time level crossing formulae that are computationally much more favorable. All approximations introduced are best for systems that have an autocorrelation close to 1. This can be interpreted as any system/process that has a high data sampling rate, or a small sampling interval.

For the facility manager that desires more control over their responses to operation of the building HVAC system, alarm systems with *no* automated subsequent control action may allow this with complete flexibility. However, for those that wish to relinquish the management of these objectives to an autonomous agent, an optimal "alarm-based control" and complaint-response algorithm can also satisfy this requirement, and is presented in Chaps. 6 and 7. The objective function is one that penalizes the actual monetary cost of energy conditioned on complaints or alarms. An optimal control law traditionally provides a control signal directly to the plant. However, the defacto standard for today's HVAC controls are PI controllers, and hence optimal control laws aren't practically implemented as a general rule. Alternatively, a setpoint generation scheme for a thermostat setting strategy might be proposed as the method by which the optimal control objectives are achieved. Developing automated thermostat setting strategies is only *one* possible way to achieve the desired

objectives, which may not necessarily be applicable for all buildings, operational modes, and HVAC system types. There may be other setpoints such as VAV (Variable-Air Volume) flow controllers, chiller setpoints, etc., for which similar optimal strategies may be derived. Although it's best to keep the optimization as general as possible so that the methods are applicable across a broad scope of setpoint strategies, the primary example in this thesis will be for a zone thermostat setting strategy.

Chapter 2

Background

2.1 Current Industrial Practice

In order to better understand the evolution of modern industrial techniques in commercial building operations, the following section provides a background outlining the framework around which future developments in mitigation of costs due to energy usage and improvements in thermal comfort are to be measured by.

2.1.1 Historical Treatment

Historically, there have been advances made in how to treat the issues of energy and thermal comfort from a fundamental and operational standpoint, independent of recent technological advances. The concepts and ideas that set a baseline for how the HVAC/facility management industry perceives these issues will be covered in this section.

Energy

It's always been important for facility managers to effectively manage and track their energy costs, which come in large part due to high consumption of electricity by commercial building HVAC

systems. In doing so, they must keep in mind some fundamental tenets of energy conservation. Some of the more important ones as stated in the California Energy Commission's (CEC) publication [39] pertaining to energy efficiency project management include the following :

- Tracking energy usage and expenditures incurred.
- Looking for "energy traps" and providing recommendations for investing in alternatives or alterations.
- Watching the energy budget and forecasting future costs.
- Looking for alternatives to traditional rate structures and energy prices in a volatile electricity market.

Concerning this last item, as a result of the deregulation of utilities in several U.S states, it is paramount to facility managers to take advantage of the potential for lower energy rates as well as different rate structures. It's also important for them to completely understand their utility bill charges. According to the CEC [39], typically, electricity bills consist of many different components. There is usually a fixed service charge, in addition to a per kW-h (kilowatt-hour) rate charge for the amount of energy consumed for the billing period. There is also a demand charge based on the peak electricity usage averaged over a short time period per billing period. For larger commercial buildings with energy consuming HVAC equipment, reactive power is also charged for increased electric transmission capacity, required for large inductive loads.

With the advent of deregulation, some other rate schedules becoming available are the time-of-use rate and real-time pricing. The time-of-use rate is one in which the demand charge and rate charge pricing varies depending on the time the electricity is consumed, typically according to season and/or time of day. With real-time pricing, customers will pay for electricity via demand and rate charges, at the exact same wholesale price as the utility pays for it. Customers may be able to obtain real-time price of utility information hours in advance of them going into effect, thereby allowing

them to anticipate higher rates and acting accordingly to reduce energy costs. This, however, will require more sophisticated hardware and software than may currently exist and are available to many facility managers. Computerization of energy usage and price tracking may become necessary to be on par with a real-time pricing rated schedule, as well as control of equipment with a DDC system and being aware of typical electricity demand on a real-time basis.

Thermal Comfort

Historically, there have been many studies pertaining to thermal comfort of building occupants. Early work involved purely empirical relations between physical variables and thermal sensation ratings from occupants in laboratory studies [76]. However, none have set the baseline for how thermal comfort is measured as much as the largely celebrated work by Fanger [16]. His PPD-PMV (Predicted Percent Dissatisfied - Predicted Mean Vote) model has been widely accepted for design and field assessment of comfort conditions, and is the industry standard for measurement of thermal comfort, as set forth in ISO (International Organization for Standardization) Standard 7730 [36], and many ASHRAE (American Society of Heating, Refrigeration and Air–Conditioning Engineers) standards as well. Gagge et. al., [26] proposed an extension of Fanger's PMV, improving the accuracy under sweating conditions.

According to Fanger, [16], the following are definitions of PPD and PMV:

PPD Predicted Percent Dissatisfied, predicts the expected fraction of a large group that will align with a subjective assessment of hot or cold above an absolute PMV level of 1.5 scale units (between slightly warm and slightly cool). It is basically an expression of the "potential complainers".

PMV Predicted Mean Vote, predicts the subjective thermal sensation rating of a large group based on 6 variables (4 environmental, 2 personal) that affect the human heat balance:

1. Activity ..... Personal

2. Clothing ..... Personal

3. Air Temperature ..... Environmental

4. Mean Radiant Temperature ..... Environmental

5. Relative Air Velocity ..... Environmental

6. Air Humidity ..... Environmental

Quantitatively, PMV is measured on the ASHRAE [5] thermal sensation scale, ranging from {-3...3} as follows:

-3 ... cold

-2 ... cool

-1 ... slightly cool

0 ... neutral

1 ... slightly warm

2 ... warm

3 ... hot

In general, approximately a 3^oC change in temperature is necessary to detect a thermal sensation vote by one unit or temperature category, according to the ASHRAE Handbook [5].

$$|\Delta T| \approx 3 \times PMV \tag{2.1}$$

where $|\Delta T|$ is the absolute deviation of the space temperature from optimal in ^oC. Fanger related PPD to PMV via the following equation:

$$PPD = 100 - 95e^{-(0.03353PMV^4 + 0.2179PMV^2)} \tag{2.2}$$

From Eqns. 2.1 and 2.2, we can derive the following relations:

PMV units	**PPD (in %)**	**$\|\Delta T\|$**	
		(in °C)	**(in °F)**
± 1	26.12%	3^oC	5.4^oF
± 0.5	10.23%	1.5^oC	2.7^oF
0	5%	0^oC	0^oF

Table 2.1: Table of Thermal Comfort Equivalences

The ASHRAE Handbook [5] guideline states that for a resting person wearing trousers and a long-sleeved shirt, thermal comfort is experienced in a still environment at $T_{opt} = 24^oC = 75.2^oF$. Here we see the use of the 2 personal comfort factors that influence the PMV in determination of the optimal thermal comfort temperature. The typical definition of dissatisfied is anyone voting more than -1, 0, or +1 PMV units in absolute value. However, if we consider the actual threshold that defines a zone of comfort around typical building occupants as opposed to the threshold of dissatisfaction, then the range of acceptable PMV values shrinks to $[-0.5, +0.5]$. We know from Table 2.1 that $\pm 0.5\ PMV\ units \Leftrightarrow |\Delta T| = 2.7^oF$. Therefore, the range defining the thermal comfort zone is $75.2^oF \pm 2.7^oF$, or $T_{opt}\epsilon\ [72.5^oF, 77.9^oF]$, where T_{opt} defines the baseline thermal comfort range based upon the ASHRAE guideline. Using T_{opt}, and combining Eqns. 2.1 and 2.2, we obtain the following relation:

$$PPD = 100 - 95e^{-(3.94\times10^{-5}(T-T_{opt})^4 + 2.56\times10^{-5}(T-T_{opt})^2)} \tag{2.3}$$

where T_{opt} and T are in ^{o}F. Figure 2.1 illustrates the qualitative nature of PPD's variation with both PMV and space temperature. There is an interesting passage in Fanger's original work, [16], where he conjectures what it means qualitatively, for example, when PMV = -0.3. We can see from Fig. 2.1 that the optimal temperature occurs at 5% PPD. Hence, it is impossible to reduce the complaint level to 0% PPD. But is PMV = -0.3 acceptable ? According to the heuristics described earlier $PMV\ \epsilon\ [-0.5, +0.5]$, it is, but people are different in their tolerance levels and there is naturally a certain variance in the thermal sensations of a group. Fanger says that any complaints, however few, are often taken as an indication that the (HVAC) system is defective, or at least badly

operated. As a result, facility operating personnel will often expend man-hours and overhead charges in response to that complaint. The corrective action is often to change the thermostat setting *simply* because of complaints of *individual* persons in a large group. So therefore these particular persons will perhaps be satisfied, but on the other hand, others may become dissatisfied. As a result even a larger number than before may complain, if an "optimal" condition existed previously. Complaints can't be altogether avoided, but they may be reduced by keeping PMV=0, or the space temperature at an optimal value.

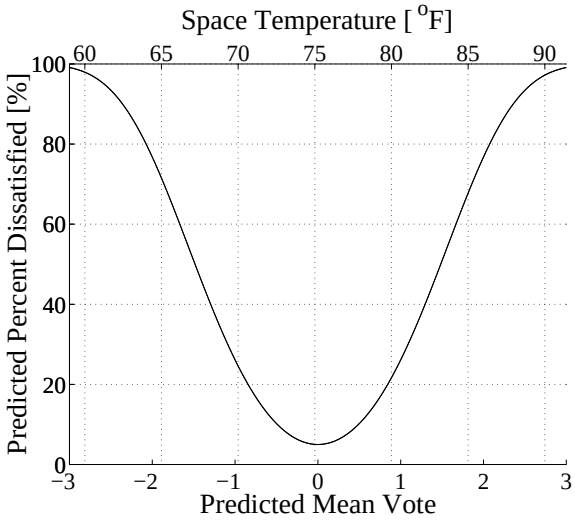

Figure 2.1: Plot of PPD vs. PMV and Space Temperature

A natural question then arises: Should facility operators change the thermostat setting in response to thermal sensation complaints ? We might consider that "warm and cold" dissatisfied levels are associated with the *average* or *majority* response of a group of people in a zone or space within a building, rather than with a single individual. Perception of temperature is influenced by several physiological and psychological factors, as well as group dynamics, which may vary from individual to individual, from time to time, and group to group. We'll discuss specific methods of responding to complaints in more detail based upon modelling these random components later in

the chapter, based primarily upon work by Federspiel [19, 22].

Information Technology

The advancement of information technology and databases over the past decade has seen quite an improvement. This has been very important in the maintenance sector of the commercial building HVAC industry. Often, tracking of maintenance labor and material costs had been difficult. Costly overruns and lack of real-time and accurate information on system status prevailed due to unsophisticated and ungainly stove-piped information repositories for facility managers having to track multiple large buildings. With the advent of newer more robust maintenance management systems that take advantage of the power of relational databases, these systems have allowed for more accurate and reliable information tracking capabilities. Fewer cost overruns, human errors and faults in HVAC systems can be attributed to lack of real-time information and unmanageable information sources.

2.1.2 Standards

Examples of prominent codes or standards pertinent to indoor environmental quality and energy usage in commercial buildings are listed and briefly described in Table 2.2.

2.1.3 Solutions and Treatment of problems

Energy

Here we list some of the specific methodologies that achieve certain energy goals, applying to what's done currently in practice that include the following:

Load shedding A strategy designed to save energy costs in peak electrical demand by picking specific periods to cut back in lighting, plug, and other non-essential loads.

Title	Organization	Primary Content
ASHRAE/ANSI Standard 55-1992, Thermal Environmental Conditions for Human Occupancy[4]	ASHRAE* ANSI†	Acceptable range for temperature, humidity, and air velocity
ISO 7730 Standard 1994 Moderate thermal environments determination of the PMV and PPD indices and specification of the conditions for thermal comfort [36]	ISO‡	Acceptable range for temperature, humidity, and air velocity
10 CFR§ Part 435 subpart A	US Dept. of Energy (Federal)	Energy efficiency in new commercial and multifamily high rise residential buildings
ASHRAE/IESNA¶ 90.1-1989[3]	Typical State Energy Code	Energy Efficient Design For New Buildings Except Low-Rise Residential Buildings
Title 24, Part 6 California's Energy Efficiency Standards for Residential and Nonresidential Buildings (meets or exceeds ASHRAE/IESNA 90.1-1989)	California Energy Commission (State)	The California Building Code establishes building energy efficiency standards for new construction (including requirements for entire new buildings, additions, alterations, and in nonresidential buildings, repairs).

*American Society of Heating, Refrigeration and Air–Conditioning Engineers
†American National Standards Institute
‡International Organization for Standardization
§Code of Federal Regulations
¶Illuminating Engineering Society of North America

Table 2.2: Standards, codes, and guidelines

Night setback A thermostat setting strategy that provides for reduction in energy usage by altering the setpoint to a higher or lower value during unoccupied periods, assuming a single day shift operation, and also depending on the season. Typically, the setting remains at a value that is commensurate with achieving thermal comfort during the occupied periods, and then at the start of unoccupied period (at night), the setpoint is adjusted to a value that will result in the usage of the least amount of energy according to the particular season. Typically in this type of strategy, the setpoint is returned to the thermal comfort value several hours prior to the beginning of the occupied period, so that the building has sufficient time to reach thermal equilibrium.

Physical Plant retrofits These are architectural and/or mechanical changes to building structure or systems, by either passive (typically architectural), or active (mechanical) means that were not originally included in the design of the building. These changes can be made to achieve grater overall energy efficiency, thermal comfort and cost savings.

Thermal Comfort

Clothing Modification Building occupants are typically able to adapt well to their surrounding environment within certain limitations. Also, due to the natural variation in the thermal comfort thresholds of individuals within a group, some building occupants dress in layers in order to accommodate for their personal thermal comfort needs. Obviously, the seasonal effect of having clothing modification as a way of coping with thermal comfort needs is evident by lighter dress in the summer season and heavier clothing in the winter season.

Physical Environment Modification Depending on the specific type of zone, building occupants may have other ways of coping with personal thermal comfort mismatches rather than clothing modification alone. Such methods include actual modification of the surrounding environment if the option is available, such as the use of natural ventilation via windows and fans, as well as personal electric space heaters, etc.

Setpoint Adjustments For zones that contain a local control point (thermostat), this is an obvious way to allow for providing thermal comfort to building occupants in a zone. The thermostat can be adjusted according the to desires of the majority of the building occupants during the occupied period. In practice, facility operators are often called on to respond to these thermal sensation complaints. They will adjust the setpoint in response to the needs of the building occupants. Sometimes building occupants themselves will "tamper" with the thermostat even if against the zone policy in order to provide themselves immediate thermal comfort. This may partly be due to the lengthy wait and/or bureaucracy of getting a timely response to the complaint, which may be caused by inefficient management of the business process. However, these inefficiencies can be due to any other number of reasons, including a poor maintenance management system, inaccurate classification of the problem causing the thermal sensation complaint, etc.

2.2 Current Research Efforts

In this section we address some of the current research efforts that attempt to improve the status quo of building operations as it pertains to HVAC systems, in addition to some of the problems and issues cited thus far.

2.2.1 Energy

Passive Solutions

There are currently many research efforts studying the use of passive or natural building designs and modifications to achieve both thermal comfort and energy objectives. Architectural and structural design in general are the subject of a study performed by Zmeureanu and Fazio [77]. They provide an example of how an integrated approach to building design determines the impact of a structural system on energy consumption. The main idea caters to integration of computer

simulation and interactive software programs that can be used to estimate the thermal performance of buildings, rather than relying on the designer's intuition. The use and integration of programs such as these into existing management portfolios would also allow for energy consumption of large commercial buildings to be estimated, and used to develop multiple alternatives for retrofitting during operation as well as in pre-commissioning design.

Although performed as a residential study, Reddy et. al. [60] present ideas which are also relevant to commercial buildings in terms of practical implementation of methodologies to save energy by using peak–load shaving. The method appeals to the intelligent use of building thermal mass, involving pre–conditioning/pre–cooling of the structural thermal capacitance and furnishings. The thermal mass is cooled during off–peak hours and slowly warmed during peak periods, hence absorbing heat that otherwise cooling air–conditioning (AC) would have to remove. This allows for the AC to be completely shut off during these periods, without affecting building occupant comfort significantly. Cool storage by means of ice or chilled water is an alternative sometimes used, but exclusively for commercial buildings.

Seem and Braun [63] present an adaptive algorithm that uses deterministic and stochastic techniques involving look–up tables with updating, in order to forecast the electrical demand in a building. It is necessary to develop forecasts of cooling loads and electrical demands for buildings that use the type of HVAC control that employs thermal storage systems, such as the one described earlier [60]. In this way, thermal storage systems can be utilized for building pre–cooling during the days identified with high cooling loads, so that peak electrical demand can be reduced.

Similar studies throughout the past decade [1, 65, 62] show that the use of passive techniques to minimize energy usage and improve thermal comfort is an ongoing effort. These are by no means the only studies, and it does not represent an exhaustive dossier of research efforts. Some of the techniques discussed include but are not limited to the use of natural ventilation, fenestration, phase-change materials, as well as several other basic components of the building fabric. One study in particular [1] considered a parametric sensitivity analysis of various building parameters including

convective heat transfer coefficients, zone temperature swings, and the thermal mass of the floor, ceiling, walls, furniture and carpeting. It was found that there are specific limitations, advantages and disadvantages to varying each of the factors. The use of passive methods does not require any energy saving actions to be implemented since they are typically design and/or construction considerations attempting to eliminate potential problems at their source. However, these methods require great attention to detail, as well as expert–level knowledge of the fundamentals of the physics and heat transfer that is involved in the complex dynamics of building thermal interactions. As such, obtaining valid quantitative descriptions of these interactions for computer simulation or other purposes is often difficult, and requires mathematical models based upon partial differential equations which may have no analytic solutions.

Active Solutions

Active solutions for the use of energy reductions can provide alternative methods in which the use of energy analyses and/or optimization become useful. There are some active solutions that deal more with the structural and architectural aspects of building design, such as ones mentioned in literature [62], including passing cool fluids or air through pathways deliberately cast into the structure of the building. However, in this section we focus mainly on the use of control algorithms and optimization–based active solutions, and choose three primary sources of research to initiate a thorough discussion. These three sources do not represent a comprehensive treatment by any means, but rather are meant to serve as a sample of the most relevant literature to our discussion.

In [11, 58, 34], optimal control is investigated as a method for achieving reduced energy usage and improved building occupant thermal comfort for thermal energy storage systems described in the previous section. Optimal controllers are hailed as the theoretical upper bound (best possible) for achieving their specified objective, minimizing some given cost function, often with the use of dynamic programming. All three papers use similar cost functions, as follows:

$$J = \sum_{k=0}^{L} R_k P_k T_s \tag{2.4}$$

where

k	=	hour of the day
L	=	Total number of hours in simulation
R_k	=	Cost of electricity at hour k
P_k	=	Plant's electric power consumption at stage k: in some cases this is a function of other variables, such as zone temperature, other uncontrolled variables, or broken down into more tangible elements such as fan, heating, and cooling power.
J	=	Cost of operating plant for the length of time given by L
T_s	=	Sampling interval (time between samples)

Each author has their own take and specific objectives in the approach they use for optimal control. Each study also has its advantages and disadvantages. For example, the earliest work by Braun [11] also imposes an additional thermal comfort constraint on the zone temperatures at each time step in the simulation. Therefore, the optimal control problem is posed as a constrained linear optimization problem over the cost of energy. It is proposed that one of the results of the optimal control solution should be the development of an efficient policy for adjusting the thermostat setting. Several alternatives are discussed and compared, including night setback, as well as dynamic building control. Dynamic building control differs from night setback in that the space temperature setpoint drifts upward during the occupied periods, as opposed to remaining constant. At the start of the unoccupied period, all equipment is turned off and the temperature floats, until just prior to the time at which occupancy is resumed, when the equipment is turned on in order to pre–cool the building. Thermostat setting policies are an important subject of Chaps. 6 and 7, so this work serves as a solid foundation.

As seen in work by Henze et. al., [34], a similar optimal control objective was posed, however it is unconstrained in the sense that there are no limitations based on thermal comfort. However, the study does provide several substantive and comparative methods for forecasting building cooling loads as well as weather information. The controller is predictive, and takes the stochastic nature

of the problem into account. A broad range of predictive methods were implemented, but all appear to achieve close to the cost savings that can be achieved by the optimal controller with perfect prediction. Even better performance can be achieved with more sophisticated algorithms, specifically the ANN (artificial neural networks).

The optimal control solution derived was different from Braun's in that a charging and discharging strategy for a generic thermal energy storage system was computed as opposed to zone temperature setpoints, so the solution caters more to that specific type of HVAC system. This controller was also compared to some other conventional control strategies typically used for this particular HVAC system type as a benchmark. Additionally, the R_k variable in the optimization shown in Eqn. 2.4 is based on several different rate structures for comparison. With the advent of the deregulation of utilities, more flexible rate pricing structure will become available, hence some fictitious sinusoidally varying rates were developed for comparison to more typical on-peak and off-peak rate structures used in a utility–regulated environment. It was found that the proposed controller can more adequately account for the complex rate structures, and in fact is more robust with respect to prediction and model quality, and has a significant performance benefit over the conventional controllers.

Finally, Price and Smith [58] show that optimal control can be used to develop strategies for shifting building mass thermal energy storage via charging and discharging, similar to the paper previously discussed [34]. However, here the quadratic cost function is slightly different in that it isolates the terms based upon the specific HVAC systems components contributing to expenditure of energy, such as fan, heating and cooling power. Additionally, it is a constrained optimization based upon thermal comfort, by using the PMV index. Similar advantageous results are achieved as in the previous two papers. They are realized through the use of optimal control, as well as other informative parametric findings concerning the amount and location of building mass and allowed limits for the PMV index. However, the rate structure used is strictly on/off–peak, and the PMV constraint does not clearly relate to dynamic building occupant thermal comfort levels, but rather

to a constant value of ± 0.5 during occupied periods.

To close this section, we note that there are several positive and advantageous results that can be derived from using optimal control with energy costs as the object of the problem formulation, with thermal comfort as one of the primary constraints. However, there are some obvious limitations as well as some restrictions in the literature as to the types of systems, operational modes, etc., which these methods have been applied to. A more general framework that considers all possibilities of building, HVAC system, operational mode types, etc. needs to be developed for optimal control to be considered as a potentially industry–enhancing breakthrough technology, even though it is relatively well-known, older, and basic in the modern controls world. This has been recognized by Braun in recent work [10]. However, the implementation of optimal control has not unanimously been endorsed, even for the attainment of more advanced objectives than local zone setpoint regulation. More simple PI controllers do a good job of this at the local level for most buildings and have modest computational memory and processing requirements. Hence, they remain the defacto standard, while the advanced objectives are often left in the hands of the facility managers, unassisted by available computational power for information consolidation. Chapters 6 and 7 will introduce an optimal control algorithm that is implemented via thermostat setting changes. The utility of this algorithm will be examined in detail, to determine the practical suitability of achieving the desired objectives. The objective function uses thermal sensation complaint statistics, while penalizing energy costs.

Because energy and complaint costs both have a monetary basis, they have an equal footing, unlike some of the other examples presented thus far, using constrained optimization, etc. The cost function might look something like Eqn. 2.5 below, which is similar to the idea presented in Martin et. al. [48].

$$J = \sum_{k=0}^{L} R_k P_k T_s + C_k \tag{2.5}$$

where

k	=	hour of the day
L	=	Total number of hours in simulation
R_k	=	Cost of electricity at hour k
P_k	=	Plant's electric power consumption at stage k: in some cases this is a function of other variables, such as zone temperature, other uncontrolled variables, or broken down into more tangible elements such as fan, heating, and cooling power.
C_k	=	Cost of complaints predicted by complaint model
J	=	Cost of operating plant for the length of time given by L
T_s	=	Sampling interval (time between samples)

2.2.2 Thermal Comfort

In this section, as in the last, specific research results and studies having relevance to the objective of this thesis are addressed. Again, they are meant to act as a representative cross–section, not as "the final word."

General Thermal Comfort

Fanger's PMV index has been the hallmark of many research efforts to develop improved methods to provide comfortable environmental conditions for building occupants. At the same time, facility designers and operators are also concerned with using the least energy possible to accomplish that purpose. Strategies have been adopted based on operating HVAC systems in the most intelligent way to make them capable of serving both purposes. In a study by Hamdi et. al. [29], a new approach to predicting the thermal sensation index of human comfort is proposed, using the six factors that the PMV index is based on. It is specifically touted as being usable for feedback control in HVAC systems. Here, fuzzy and neuro-fuzzy methods were used for the development of classification functions. An example was presented with the practical use of heuristic rule-based methods. Fuzzy methods are seemingly a very nice fit given the qualitative and linguistic nature of how the PMV index is constructed (cold, cool, slightly cool ...).

The main reason for developing this method is to solve a problem with computing the

PMV directly from the six factors. Fanger gave a very complex nonlinear equation relating PMV explicitly to these six variables. It is much too complex and is not suitable for use in real-time application with on-line adaptive calibration and feedback control of HVAC systems, due to the iterative nature of the computations required to arrive at a solution. Previous work [66, 17] has shown that there may exist ways to simplify the PMV equations so that they are more easily manageable for computation. The main simplifications in these studies were linearization assumptions allowing for explicit computation. However, the drawbacks involved linear parametrization problems as well as some non-linear effects that became important when considering all variables.

In summary, Hamdi et. al. [29] provide a step in the right direction, because none of the above simplifications were made, and no iterations were required to arrive at a result for obtaining the PMV. Additionally, the results for different combinations of the factors involved in the PMV model are quite near the computed actual PMV values for the same combinations using Fanger's formulas. The idea of using fuzzy logic to obtain these results is an insightful one, however the practicality and actual testing and/or simulation of this new fuzzy-based PMV index in an actual feedback control system was not presented. This new algorithm also seems to be heavily based upon the use of subjective designer rule–based heuristics. Additionally, all six variables *must* be qualified with some linguistic modifier in order for this algorithm to compute the PMV.

It is difficult to qualify the personal comfort variables for a group of building occupants in a zone, and probably a better idea to quantify the environmental factors than qualify them. Unfortunately, there is still no good way to measure PMV for use in a control system, to obtain an adequate representation of the thermal comfort level of the majority of the occupants. However, we may be able to use the building occupants as a source of information to qualify and quantify their *own* comfort levels, rather then measuring it via the factors of the PMV and assuming that it's an accurate representation. Thermal sensation complaint data exists within the databases of maintenance management systems and is ubiquitous, "free" data. It also provides an indication of comfort *directly related* to the human condition. The database or the building occupants themselves

can be thought of as "sensors" that capture the essence of the informational content of the existing comfort level. Therefore, unsolicited building occupant feedback can be thought of as information that may suitable for use in a feedback control system. Although it is not directly measurable, or quantifiable with the PMV index, there are ways to shape and consolidate this data.

Solicited Occupant Feedback

Questionnaires have been used in research studies [74, 42] in order to solicit occupant feedback concerning general and specific perceptions of their immediate surrounding environment. The key factors that building occupants were asked about or find most important can be condensed down to the following list (not necessarily listed in order of importance):

1. Thermal Comfort (temperature)
2. Air Quality (ventilation and health)
3. Lighting
4. Control of Environment
5. Space and Privacy

Although facility engineers may put effort into designing and measuring environmental variables to provide comfortable conditions, rarely are the occupants themselves polled as to how satisfied they are with their building and its HVAC services. Therefore, the use of occupant feedback as a viable information source is somewhat under-utilized. However, the use of occupant feedback is not completely without its limitations. In a perfect world building occupants could be modelled as virtual sensors. However, their behavior is not at all predictable due to learned responses, as well as sociological and cultural impacts of obtaining information from them. In fact, fluctuations in occupants' comfort levels can be related to several different confounding factors, including actions or inactions of building managers. Yet it still holds that the majority of information received

from building occupants is typically quite beneficial, and can be integrated with other sophisticated monitoring techniques and building systems for the overall goals of improved occupant thermal comfort and satisfaction.

A study by Fountain et. al. [24] uses the technique of soliciting feedback in order to gain knowledge about thermal comfort for short–term occupancy in hotels. This was done to provide a foundation for the development of a new thermostat control system that implements microprocessor–based logic, addressing the needs of temporary occupancy while reducing energy consumption. This study provides an excellent example of how questionnaires and solicited occupant feedback can be used pragmatically. The use of the questionnaires were beneficial in the design of the control system by finding the thermal expectations and preferences of the building occupants. Some of these included the need for user–friendly thermostat system controls, annoyance with slow HVAC system response to commands, poor air quality, uncomfortable thermal conditions, etc. This feedback aided in the design of the new control system, which accounted for this information, and used occupancy sensors, and other techniques such as controlled temperature drift, etc. in order to both provide for the dual objective of thermal comfort and reduced energy usage.

Unsolicited Occupant Feedback

Preliminary treatment of the impact, statistical analysis and conjectured use of *unsolicited* occupant feedback of thermal comfort conditions was first presented by Federspiel [18]. The study presents a comprehensive examination of complaints taken from a large empirical dataset from actual operating buildings. This includes measurements of frequencies and statistics of temperatures, response times, and actions taken in association with complaints found in logs. The author stated that prior to the development of operational policies for managing unsolicited complaint costs, a basic yet thorough understanding and presentation of unsolicited complaint behavior needs to be completed. Many key aspects of building operations that may cause the conditions leading to complaints and general diagnostic lapses had not been well documented prior to the study. The

most interesting findings of this initial study were the following :

1. Thermal sensation (hot or cold) complaints are the most common type of unsolicited environmental complaints in buildings.

2. Thermal sensation (hot or cold) complaints are mainly caused by either control system performance problems or faults in the HVAC system, as opposed to building occupant preferences.

3. Most actions taken in response to thermal sensation complaints involve adjusting a control system setting of some sort.

4. There is great potential for reducing HVAC maintenance labor costs by reducing the frequency of hot and cold complaints.

In a follow–up paper, [19] Federspiel presents a statistical model based upon level–crossing theory to develop a way to predict the frequency of hot and cold complaints. Data from an actual commercial building was measured and used to develop estimates of the model's parameters. The parameters of interest are the frequency (ν_T), mean (μ_T), standard deviation (σ_T), and standard deviation of the rate of change ($\sigma_{\dot{T}}$) of temperatures at which building occupants complain of being too hot or too cold. This model is of great importance to this thesis, and provides us with the static equivalent to a dynamic extension of it in future work. [46, 47]. Theoretically, we can think of there being 5 processes in this model. The main process is the building temperature trajectory, and each hot and cold complaint level has 2 processes associated with it as follows:

Complaint Level or Process A continuous trajectory representing the fictitious statistically–based temperature level at which that the majority of building occupants in a zone complain of being too hot or cold.

Complaint Events A discrete random sequence which represents the exact instants in time at which the building occupants complain of being to hot or cold. This process coincides with the continuous building temperature trajectory at these particular instants.

Figure 2.2 is provided to aid in illustrating the concepts stated above.

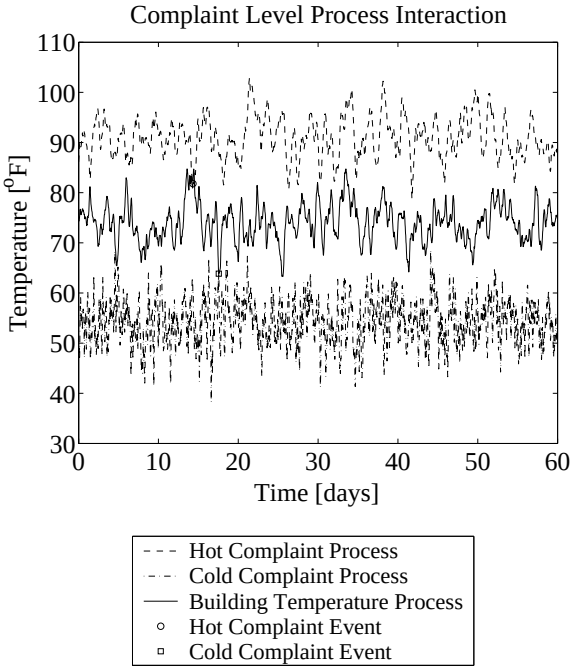

Figure 2.2: Plot of Complaint Process Interaction with Building Temperature

Essentially, the continuous complaint level statistics are unknown model parameters which cannot be measured and are shown in Table 2.3 (capital subscripts used for continuous levels).

	Mean of T*	**Standard Deviation of T**	**Standard Deviation of $\dot{T}$**
Hot	μ_{T_H}	σ_{T_H}	$\sigma_{\dot{T}_H}$
Cold	μ_{T_L}	σ_{T_L}	$\sigma_{\dot{T}_L}$

*Refers to an arbitrary complaint temperature level

Table 2.3: Table of Non-Measurable Complaint Model Parameters

Estimates of the non-measurable model parameters in Table 2.3 are mathematically based upon the measurable statistics shown in Table 2.4 (lower case subscripts used for discrete event processes and capital subscripts used for continuous levels). Assuming that building temperatures are uncorrelated with, or independent of the high and low complaint levels, the formulas relating

the parameters of the model listed in Table 2.3 explicitly to the data measurements listed in Table 2.4 are shown in Eqns. 2.6- 2.11.

	Frequency	**Mean of T**	**Standard Deviation of T**	**Standard Deviation of $\dot{T}$**
Hot	ν_h	μ_{T_h}	σ_{T_h}	
Cold	ν_l	μ_{T_l}	σ_{T_l}	
Bldg		μ_{T_B}	σ_{T_B}	$\sigma_{\dot{T}_B}$

Table 2.4: Table of Measurable Statistics

$$\mu_{T_H} = \frac{\mu_{T_h}\sigma^2_{T_B} - \mu_{T_B}\sigma^2_{T_h}}{\sigma^2_{T_B} - \sigma^2_{T_h}} \tag{2.6}$$

$$\mu_{T_L} = \frac{\mu_{T_l}\sigma^2_{T_B} - \mu_{T_B}\sigma^2_{T_l}}{\sigma^2_{T_B} - \sigma^2_{T_l}} \tag{2.7}$$

$$\sigma_{T_H} = \sqrt{\frac{\sigma^2_{T_h}\sigma^2_{T_B}}{\sigma^2_{T_B} - \sigma^2_{T_h}}} \tag{2.8}$$

$$\sigma_{T_L} = \sqrt{\frac{\sigma^2_{T_l}\sigma^2_{T_B}}{\sigma^2_{T_B} - \sigma^2_{T_l}}} \tag{2.9}$$

$$\sigma_{\dot{T}_H} = \sqrt{\frac{4\pi^2\nu_h^2\sigma^4_{T_B}}{\sigma^2_{T_B} - \sigma^2_{T_h}} e^{\frac{(\mu_{T_B}-\mu_{T_h})^2}{\sigma^2_{T_B}-\sigma^2_{T_h}}} - \sigma^2_{\dot{T}_B}} \tag{2.10}$$

$$\sigma_{\dot{T}_L} = \sqrt{\frac{4\pi^2\nu_l^2\sigma^4_{T_B}}{\sigma^2_{T_B} - \sigma^2_{T_l}} e^{\frac{(\mu_{T_B}-\mu_{T_l})^2}{\sigma^2_{T_B}-\sigma^2_{T_l}}} - \sigma^2_{\dot{T}_B}} \tag{2.11}$$

The stated independence assumption, made in Federspiel's study, [19], is valid if the coping behaviors of building occupants do not vary with building temperature. The performance of the temperature control system can be characterized by the measured building temperature statistics. The mathematical model relates them to complaint levels. Examples of how the model might be used in optimization of the operational performance of buildings was presented. This involved a balanced tradeoff between energy usage and thermal comfort need. Additionally, an economic analysis predicted significant cost savings. An example of this is presented in other work [47, 46], where computer simulation methods were used to compare two different strategies for responding to unsolicited thermal sensation complaints in buildings. The modelling parameters for Federspiel's

original work [19] have also been validated in a follow–up study, [22, 21]. These re–calibrated parameters provide the basis for the optimal control and alarm results in Chapters 6 and 7. The optimal alarm results presented in Chap. 3 are based upon the original model parameters from [19].

2.2.3 Commercial Building HVAC Systems

The design, operation and maintenance of commercial building HVAC systems has seen most innovation in recent years within the more practical aspects of the field. However, there is certainly room for innovative ideas and theories to be developed in order to improve overall HVAC system advancement and utilization. There are some recently published papers [51, 64], which address some of the new technology that is available in current practice, and expresses the potential for the use of available information sources with this new technology using fundamental ideas.

In [51], an introduction to DDC shows that microprocessors, memory, user interfaces, communications, and power supplies are now ubiquitous in the HVAC industry, and are making continuous advancements. These components are part of modern building automation systems and they can and should drive the future of applying more sophisticated and advanced control algorithms. In this way they can lead to higher levels of integrated building operational performance, satisfying thermal comfort needs as well as saving energy.

Seem et. al. [64] develop a new "control performance monitor", which implements a basic EWMA–based algorithm to compress and integrate large amounts of data. EWMA is the exponentially–weighted moving average, which can simply implemented as a digital filter as follows:

$$\zeta_{k+1} = (1-\lambda)\zeta_k + \lambda x_k \tag{2.12}$$

$$z_k = (1-\lambda)\zeta_k + \lambda x_k \tag{2.13}$$

where

ζ_k	=	Internal state of digital filter at discrete instant k
x_k	=	Input to filter (measured data)
z_k	=	Output of filter (EWMA value)

The EWMA parameter, λ, also known as "the forgetting factor," places more weight on recent values of the measured variable of interest than past values when assigned a large value, and less weight on recent values when assigned a smaller value. Since modern building automation systems have such a large capacity for storing data, there is sometimes an information overload at the end–user point of access. Hence there seems to be an inordinate amount of data, but no clear theories or methods that have been developed to make sense of all of it in a simple fashion that makes facility operators' and managers jobs simpler rather than more complex. Some of the variables that are monitored can be difficult to interpret. One example provided is that of control alarms. With the ability to perform trending and produce nice graphs with modern automation systems, exceedances of critical HVAC system thresholds are easily detectable. However, the limits themselves are often set by the manufacturer, and on–site facility operators may not have the same training or expertise to determine if the alarm limits are dubious. If the limits are too wide, faults may go undetected, and if limits are too narrow, false alarms may result.

Seem's treatment of condensing data by using the EWMA method will allow building operators to identify HVAC equipment that is wearing prematurely due to excessive movement. This is done by measuring the EWMA duty cycle of actuators in the HVAC system. Other EWMA indicators include control system error, VAV box airflow error, whose out–of–limit values might indicate either unstable or poorly tuned PI controllers. Basic indicators such as these may also help to aid facility operators to prioritize maintenance for HVAC system components having the highest normalized EWMA values. Since there are different EWMA indicators, response or investigation of problems causing these out–of–limit values can be based upon the current management policy. If management views uncomfortable occupants as more urgent than flow errors in a VAV box, then the rules for such response are purely left to their discretion. The EWMA for control system temperature

error may provide a potential indication of uncomfortable conditions for occupants, and the EWMA for flow errors in a VAV box might be indicative of a blockage or worn component. The information is available for interpretation and the response of those operating the system can be completely decoupled from it.

These performance indices were also conjectured to be potentially useful in a real-time fault detection system, aided by comparison of stored data in an expert system database. Additionally, diagnostic systems can be developed from pattern use and trending of the performance indicators to identify causes of faults, helping to create and populate the expert system database. Seem's "control performance monitor" idea is an excellent way of consolidating information for the express purpose of addressing the problems cited at the very beginning of this thesis. Practical implementation is not a roadblock since the algorithm requires so little memory. However, the scope of indicators used might be broadened to include such metrics as the predicted complaint rate.

2.2.4 Hybrid Issues

Attempts to link together several common problems with the control of HVAC systems to the objectives of improved occupant thermal comfort and reduced energy usage from an enterprise–level standpoint has seen some interesting treatment in research over the past decade. Fountain et. al. [24] used varied information sources including solicited occupant feedback, occupancy sensors, and other techniques such as controlled drift of thermostat setpoints , etc. in order to provide for the attainment of these objectives. However, heuristic methods were used as part of developing a subjective rule-base in the control system. As is well–known, there is no known analytic approach for a theoretical derivation and analysis of a control system using such rules. Miriel and Fermanel [50], develop similar rules in order to deal with improved performance and thermal comfort, for a specific type of HVAC system (gas boiler regulation). Here thermostat setpoint policies for residential use are derived, based on a more formal appeal to fuzzy logic. The thermostat essentially "learns" the characteristics of the HVAC system, and as such provides a satisfactory comparison to a classical

controller, both in simulation and experimentally in terms of energy reduction and thermal comfort.

The most comprehensive treatment of these issues from an "enterprise" level, is shown in [67, 71, 72]. Shoureshi et. al. [67] take the first look at the integration of multiple methodologies to achieve the objectives of occupant comfort, minimum energy, as well as maintenance cost. The methods used are fuzzy logic, optimization and optimal control, and is implemented from a hierarchical , enterprise–level standpoint. The author recognizes the need to develop a control system that can 'learn' the characteristics of the building and determine the best control strategy regardless of building type, HVAC system type, disturbances, etc., accounting for time-varying effects. In this way the controller designed is a 'supervisory controller' of sorts, and can develop new strategies to achieve the desired objectives. A theoretical treatment is given to the derivation of the algorithm that is essentially a fuzzy control system implemented hierarchically. Standard LQ (linear quadratic) optimal control methods are integrated with the developments of fuzzy logic formulation in matrix/state–space realization. Now there are some disadvantages due to the introduction of the fuzzy control, which has to do with the fact that a true optimal solution in the theoretical sense cannot be attained. This is due to a potentially long optimization computing time requirement, which is on the same order of magnitude as when the result of the optimization is needed. However, this may be less of a problem now with improved processing and memory capabilities.

The main goal of the supervisory controller, however, is cost optimization. The results provide setpoints for different HVAC control subsystems. Again, motion detection via sensors similar to the presentation by Fountain et. al. [24] is proposed. System identification is also another theoretical construct which becomes useful in the supervisor. The author creates some fictitious cost functions, taking into account similar quantities as in [34, 11] for the energy penalty. Furthermore, he develops a comfort penalty function based mathematically upon linear combinations of sigmoid functions of the form:

$$y(x) = \frac{1}{1 + e^{-\alpha x}} \tag{2.14}$$

This comfort penalty expense becomes part of the optimization itself, as opposed to being a constraint. The cost associated with lost of productivity is how the penalty is measured, and it is parameterized by deviation from "optimal" setpoints, as well as α, which characterizes this cost associated with limiting temperature deadband effects. An example of such a comfort penalty function is show in Figure 2.3.

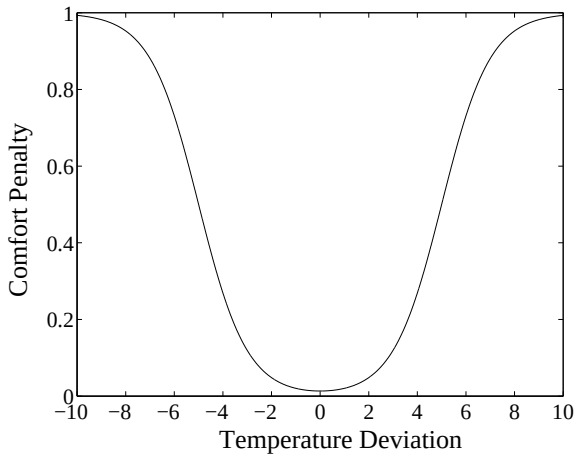

Figure 2.3: Sample Thermal Comfort Penalty Function

This treatment is very comprehensive and does a good job of generalizing the plant in question, rather than having application only to a particular type of system. However, the manner in which thermal comfort is penalized is devoid of using any measurements of actual thermal comfort of building occupants, and instead relies on a fictitious extrapolation of the most commonly known factors affecting thermal comfort (namely temperature). Also notice that with proper choice of α the comfort penalty function resembles Figure 2.1, which is simply the PPD–PMV relation.

Follow-up work based on this original study was performed by Teeter and Chow [71], in which the incorporation of artificial neural network and fuzzy logic methods were used for system identification and control of an HVAC system. The authors stress the importance of adaptive and learning techniques to be integrated with optimal control and other advanced techniques for the HVAC field. The results indicate that ANN's demonstrate the capacity to learn changes in plant

dynamics, and are able to accurately predict future behavior. ANN's in general have been found to exhibit an excellent propensity for learning HVAC plant models, and have been considered in many other studies for comparison to other techniques, some of which are based on first principles. Henze et. al. [34] as well as Krieder and Haberl [40], Haberl and Thamilseran [28] show that the performance of ANN's are superior in such comparisons. In the last two of these studies, the accuracy of calibrated models for prediction of energy used was reviewed in a "competition" to determine which of the contestants could provide the best match to an unknown validation data set from a building, given a training set. ANN's and/or statistical models always came in first.

Finally, Breeman and DeVries [72] provide a slightly different approach, although there are some similarities in terms of using a hierarchical control structure. The key idea is that complex problems for which standard solutions are not available can be solved by integrating existing solutions. In doing so, the concept of autonomous agents in the field of AI (artificial intelligence) is employed. Each agent has its own purpose, and combining them together by highlighting all dependencies gives a supervisory controller approach to solving the problem. Some of the methods that are used by individual agents include finite state machine logic, ANN's, basic local PI and PID (Proportional-Integral-Derivative) controllers, etc. The objectives that they achieve are set-point generation, adaptive tuning, and basic thermostat operation to attain user comfort, minimal switching, and robust control performance of the HVAC system.

The central idea is based upon a divide and conquer bottom–up approach to developing a solution, and *structuring* the problems coherently. Potentially, this whole framework might be suitable for analysis and synthesis by using methods still being developed in the area of hybrid control theory. This theory deals with both continuous and discrete event systems operating in unison. The objectives discussed earlier in previous research [67, 47, 64] may be good candidates for this independent agent–supervisory control methodology, and hybrid control theory may be used for stability analysis and supervisory control synthesis.

Although much of the preceding two sections have been dedicated to the treatment of HVAC

system research at an "enterprise" or hybrid level, an important research idea introduced by Seem et. al. [64] was also presented. It was discussed how exceedances of critical HVAC system thresholds are detectable, but the alarm limits often have dubious settings. In Chapter 3 and 7 of this thesis, we provide the theory and design guidelines for implementing an *optimal* alarm system, which elicits the fewest number of false alarms for a fixed detection probability. Although the alarm system is intended for anticipation and warning of future thermal sensation complaints, the fundamental idea presented can easily be extended to an HVAC system and relevant monitored processes as proposed by Seem et. al. [64]. In fact, it is possible that the methods presented can be used for fault detection of a general nature, contingent on the fact that a critical level/threshold of some sort exists, with respect to some random process.

Chapter 3

Optimal Alarm

3.1 Introduction

This chapter introduces the theory and methodology that combines the practical appeal of Kalman prediction techniques with level-crossing theory and optimal alarm system design. A comprehensive demonstration of practical application for the design of optimal alarm systems has been covered in the literature [69, 70]. However, the theory and practice for optimal alarms has also seen modest coverage by other authors as well [8, 15, 45, 59]. The latter is by no means a comprehensive list, but illustrates a cross-section of the primary authors responsible for introducing optimal alarm systems in a classical and practical sense. It was shown [69, 70] that an optimal alarm system can be derived based upon a likelihood ratio criterion via the Neyman-Pearson lemma. This lemma allows us to design an optimal alarm system that will elicit the fewest possible false alarms for a fixed detection probability. The likelihood ratio can be written in terms of an event's conditional probability, resulting in an intuitively appealing inequality: $P(C|\mathcal{D}) \geq P_b$. This basically says "give alarm when we predict that the probability of the event, C, exceeds the level P_b." Here, $\mathcal{D}$ represents data being conditioned on, and P_b represents some optimally chosen border or threshold probability with respect to a relevant alarm system metric.

There may be several applications for the theory and methodology introduced here. However, we develop the theory in the section that follows based only on a particular example, without generalization to all possible types of level-crossing events, C. While not general, this example is still unique and has not been covered in the literature before. Optimal alarm literature focuses on examples for which there is only a single level-crossing event.

Now, while it is true that the theory developed in the following section is based on a specific example, this example does have the potential for moderate generalizations. The two generalizations that need to be made are fairly innocuous. One has to do with the definition of the level crossing event itself. For our particular application, we're actually interested in any number of thermal sensation complaints, or level-crossing events that may occur over a specified time period. This will result in a complicated multi-dimensional event over a time period spanning many time slices, rather than just a single level-crossing with two adjacent time slices. This time period can be defined in multiple ways. Within the building operations domain, shift-based time-of-day complaint rates and energy usage fluctuate in a predictable manner. Therefore, it is common to expect a peak in the complaint rate during a specific time of the day, specifically during the morning, or what's termed as the "arrival complaint period." The arrival complaint phenomenon was also hypothesized by Federspiel [21] as being due to a naturally high metabolic rate during this period. Hence we can look at breaking down the periods of interest into two distinct timeframes, as itemized below.

Arrival and operating complaints Prior to the start of the beginning of the day (eg. 8 am), we want to predict an arrival complaint (i.e. an exceedance), and all remaining operating complaints (i.e upcrossings) for a sliding window of time of fixed length. Note that an arrival complaint has no restriction on happening at a particular time (i.e. late arrivals are allowed).

Operating complaints only Following the start of the day (i.e. conceivably after the first arrival complaint), we want to predict all operating complaints any point after the beginning of the day, for a sliding window of time of fixed length.

The difference between exceedances and upcrossings is as follows:

Exceedance A one-dimensional level-crossing event, $\{x_k > L\}$, where L is some critical threshold level exceeded by a process whose value at time k is x_k

Upcrossing A two-dimensional level crossing event, $\{x_k < L, x_{k+1} > L\}$

The second generalization that needs to be made is on the basis of the spectral content of the process and the critical level itself. For the first example considered, and the one which the theory that follows is based upon, the critical level, L, is assumed to be a fixed, static value. In our application, the critical value is not fixed, in fact, there are two stochastic critical levels, one for hot complaints, and another for cold. These levels represent the temperature at which a group of occupants in a zone would complain if too hot or cold. They are somewhat artificial, because such temperatures cannot be measured continuously. However, when complaints do occur, the temperatures can be measured and stored in a maintenance management database. Therefore the statistics of these levels can be computed from this database repository [19], and used to generate a very simple 2^{nd}-order model, whose output represents these complaint levels of interest. Both levels can be characterized as simple 2^{nd}-order stationary, Gaussian processes, as well as the building temperature process in question [19]. This generalization is perfectly acceptable, as the idea of the critical level itself being modelled as a stationary, Gaussian process was also recognized by Svensson as a potential candidate for study [69], p.93. The primary example to be used in this chapter is the hot complaint level, and its interaction with the controlled process of interest (building or space/zone temperature).

The most important aspect of this chapter covers theoretical derivations and comparisons of different approximations for these complicated multi-dimensional events. We need to characterize these events in order to determine the alarm region resulting from the conditional inequality: $P(C|\mathcal{D}) \geq P_b$. It is necessary to find these alarm regions in order to design the alarm system. Alarm design requires computation of the Type I/II error probabilities for various border probabilities (P_b).

Type I/II error probabilities are the probabilities of false alarm and missed detection, respectively. In alarm design, we want to find the value of P_b that provides the best tradeoff between Type I and Type II errors, or a similar alarm system metric.

For complicated multi-dimensional events, it's not feasible to obtain the *exact* alarm system metrics by means of numerical integration. Instead, one may perform simulations to obtain an estimate of the exact Type I/II error probabilities. These simulation-based statistics have well known estimation error properties. They are obtained by running a Kalman predictor, and counting the number of correct/false alarms and correct/missed detections, for a long enough time until their relative frequencies converge to limiting probability values. However, with the aid of some approximations, we can perform numerical integrations of complex integrals, and can avoid otherwise often very time and computationally intensive simulation runs.

One might question the merit of using such an unavoidably costly and computationally intensive technique to design an alarm system, when a simpler one might do. There are several different types of alarm systems, ranging in frequency of use and expense. At one end, we have alarm systems that require little economic investment other than the accumulation of man-hours of experiential knowledge. These systems, although quite inexpensive and often lacking in sophistication, tend to be the most ubiquitous in engineering systems. For the most part, they tend to the job that is required of them: to give alarm for prevention of catastrophic events. Here, false alarms would cause loss of production and capital due to system downtime from the inevitable system shutdown as a result of the alarm. However, missed detections may cause damage, complete system destruction or loss of life, where the costs are immeasurable. Since the events that need to be predicted by these more simple alarm systems are extremely critical, Type I errors (probability of false alarms) are obviously much less costly than Type II errors (probability of missed detections). Therefore, simple alarm systems are designed for Type II error avoidance. On the other end, engineering systems that predict events of a less severe nature may require a more subtle balance between Type I and Type II errors. In this case, more sophisticated alarm systems may be of interest.

The following list provides a variety of alarm systems, ranging from the most simple one to more sophisticated ones alluded to above.

Simple Alarm System Typically there is no cost for this type of alarm system, and it is based upon the knowledge and experience of the users, i.e. heuristics. The basic idea is that certain tolerances are chosen apriori, via trial and error, to provide a window of operation within which the process should be constrained.

Predictive Alarm System An alarm system that uses some sort of predictive method is often called a naïve alarm system [45, 69, 70]. Here, a predicted process value would trigger an alarm if it exceeds some fixed, pre-selected alarm threshold. However, even though the predictor may be optimal in the least-squares sense, the alarm system would not be optimal in the sense that it triggers the fewest false alarms for a fixed detection probability.

Optimal Alarm Systems Described before, an optimal alarm system is derived based upon a likelihood ratio criterion via the Neyman-Pearson lemma. The resulting optimal alarm system requires the use of predictive techniques to elicit the fewest possible false alarms for a fixed detection probability. As stated earlier, there are several approximations which may be used in cases when computing the exact optimal alarm system is infeasible. These approximations are listed below.

1) Two-Dimensional Alarm System Also called a semi-naïve alarm system in the literature [69, 70], it uses the idea of optimal alarm. However, the two-dimensional alarm region is approximated with asymptotes to "rectangularize" it, making the region of integration much easier to define.

2) Multiple Sub-Interval Alarm System We may also use the union of disjoint sub-intervals to approximate the exact alarm region. This approximation was again a recommendation of Svensson [69]. Exact alarm regions often involve complicated multi-dimensional events. Therefore, aggregating less complicated 2-dimensional alarm regions reduces the

computational load and increases mathematical tractability. Each sub-interval can be approximated with asymptotes, again making the regions of integration much easier to define.

3) Multi-Dimensional Alarm System The exact alarm system metrics cannot feasibly be obtained for complicated multi-dimensional events by means of numerical integration. However, a newly proposed approximate alarm region of integration can be defined as a tight bound on the exact region via the unions and/or intersections of planes. This approximation forms a semi-infinite hyper-rectangular region in multi-dimensional space, which approximates the region.

When Svensson first introduced examples for the practical application of optimal prediction and optimal alarm design, one was a basic 2^{nd} order model. However, its spectral content is quite different than 2^{nd}-order models based on complaint statistics described in Federspiel's work, [19]. The latter model is the one of practical interest to us. Since the dynamics of Svensson's 2^{nd}-order model are much faster, finding the exact Type I/II error probabilities for the exact alarm region will take much less simulation time. Therefore, this model will be used as a baseline comparison example. It will also serve as the primary example for developing all theory that follows, for operating complaints only. The theory can easily be extended/generalized to the other cases of interest. The other examples are provided in the following list:

1. Svensson's 2^{nd}-order model, fixed critical threshold, both arrival and operating complaints
2. Federspiel's 2^{nd}-order models, random threshold, operating complaints only
3. Federspiel's 2^{nd}-order models, random threshold, both arrival and operating complaints

3.2 Theory

3.2.1 The Exact Alarm Condition

We know that the only way to obtain enough information for designing an optimal alarm system and exact computation of Type I/II error probabilities is by simulating the exact conditions for alarm. The condition for alarm, for operating complaints only, is based upon the null hypothesis shown in Eqn. 3.1.

$$
\begin{aligned}
\mathcal{H}_0 &: \text{At least one complaint during normal building operating hours} \qquad (3.1)\\
\mathcal{H}_0 &: (\mathbb{X} \in \Omega_{C_{exact}} \subset \mathbb{R}^{m+1})\\
m+1 &= \text{dimension of hypothesis-based event}\\
m &= \text{number of steps in prediction window}\\
\Omega_{C_{exact}} &= \text{m+1-dimensional event region}\\
\mathbb{X} &= \begin{bmatrix} x_k \\ x_{k+1} \\ \vdots \\ x_{k+m} \end{bmatrix}\\
C_{exact} &= \{\mathbb{X} \in \Omega_{C_{exact}} \subset \mathbb{R}^{m+1}\}\\
&= \neg\{x_k < L, x_{k+1} < L, \ldots, x_j < L, \ldots, x_{k+m} < L\} \setminus \{x_k > L\}\\
&= \overbrace{\neg}^{\text{not}} \underbrace{\{x_k < L, x_{k+1} < L, \ldots, x_j < L, \ldots, x_{k+m} < L\}}_{\text{no complaints}} \overbrace{\setminus}^{\text{take away}} \underbrace{\{x_k > L\}}_{\text{arrival complaints}}\\
&\quad x_j < L,\ \forall j \in \{k+2, \ldots, k+m-1\}\\
L &= \text{Fixed critical threshold level}\\
x_{k+i} &= \text{Output and/or future values of a stationary Gaussian random sequence}
\end{aligned}
$$

In our specific case above, the exact condition for alarm results in the alarm condition

inequality shown in Eqn. 3.2, via the Neyman-Pearson lemma [69, 70].

$$\begin{aligned} P(x_k < L|\mathcal{D}) - P(x_k < L, x_{k+1} < L, \ldots, x_{k+m} < L|\mathcal{D}) &\geq P_b \\ \mathcal{D} &= \{x_0, \ldots, x_k\} \end{aligned} \tag{3.2}$$

Although we won't cover the theoretical derivation for the case of both arrival and operating complaints in detail here, note that the alarm condition for both arrival and operating complaints is not too different than the one shown in Eqns. 3.1-3.2:

$$\begin{aligned} C_{exact} &= \{\mathbb{X} \in \Omega_{C_{exact}} \subset \mathbb{R}^{m+1}\} \\ &= \neg\{x_{k+d} < L, x_{k+d+1} < L, \ldots, x_{j+d} < L, \ldots, x_{k+m+d} < L\} \\ &= \overbrace{\neg}^{\text{not}} \underbrace{\{x_{k+d} < L, x_{k+d+1} < L, \ldots, x_{j+d} < L, \ldots, x_{k+m+d} < L\}}_{\text{no complaints}} \\ &\quad x_j < L, \ \forall j \in \{k+d+2, \ldots, k+d+m-1\} \\ \text{where } d &= \text{number of steps prior to the start of the beginning of the day} \end{aligned}$$

A similar alarm condition inequality results:

$$1 - P(x_{k+d} < L, x_{k+d+1} < L, \ldots, x_{k+d+m} < L|x_0, \ldots, x_k) \geq P_b$$

Let's now return to the exact alarm condition for **operating complaints only**:

$$C_{exact} = \{\mathbb{X} \in \Omega_{C_{exact}} \subset \mathbb{R}^{m+1}\} = \neg\{x_k < L, x_{k+1} < L, \ldots, x_{k+m} < L\} \setminus \{x_k > L\}$$

The exact alarm condition is as follows:

$$
\begin{aligned}
A_{exact} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{exact}} \subset \mathbb{R}^{m+1}\} \\
&= \{\hat{\mathbb{X}} : P(C_{exact}|\mathcal{D}) \geq P_b\} \\
&= \{\hat{\mathbb{X}} : P(x_k < L|\mathcal{D}) - P(x_k < L, x_{k+1} < L, \ldots, x_{k+m} < L|\mathcal{D}) \geq P_b\}
\end{aligned}
$$

where:

$$
\mathbb{X} = \begin{bmatrix} x_k \\ x_{k+1} \\ \vdots \\ x_{k+m} \end{bmatrix}, \; \hat{\mathbb{X}} = E[\mathbb{X}|\mathcal{D}] = \begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+1|k} \\ \vdots \\ \hat{x}_{k+m|k} \end{bmatrix}
$$

and:

$$
\mathcal{D} = \{x_0, \ldots, x_k\}
$$

Hence it is easy to write the formulae for correct alarms/detections:

$$
\text{Let } \mathbf{x} = \begin{bmatrix} \mathbb{X} \\ \hat{\mathbb{X}} \end{bmatrix} \tag{3.3}
$$

$$
\begin{aligned}
\text{Correct Alarm: } P(C_{exact}|A_{exact}) &= \frac{P(C_{exact}, A_{exact})}{P(A_{exact})} \\
&= \frac{\int_{\Omega_{C_{exact}}} \int_{\Omega_{A_{exact}}} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \; d\mathbf{x}}{\int_{\Omega_{A_{exact}}} \mathcal{N}(\hat{\mathbb{X}}; \mu_{\hat{\mathbb{X}}}, \Sigma_{\hat{\mathbb{X}}}) \; d\hat{\mathbb{X}}}
\end{aligned} \tag{3.4}
$$

$$
\begin{aligned}
\text{Correct Detection: } P(A_{exact}|C_{exact}) &= \frac{P(C_{exact}, A_{exact})}{P(C_{exact})} \\
&= \frac{\int_{\Omega_{C_{exact}}} \int_{\Omega_{A_{exact}}} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \; d\mathbf{x}}{\int_{\Omega_{C_{exact}}} \mathcal{N}(\mathbb{X}; \mu_{\mathbb{X}}, \Sigma_{\mathbb{X}}) \; d\mathbb{X}}
\end{aligned} \tag{3.5}
$$

We know that computing these integrals, specifically, the alarm regions, A_{exact}, is an intractable problem. Therefore we must simulate to obtain estimates of the Type I/II error probabilities, in lieu of storing points along the border of the alarm region in an m-dimensional grid.

But instead of using the "counting" method via simulation described earlier, we can compute their tractable, and much less computationally intensive approximations.

3.2.2 Kalman Filtering and Prediction

Before attempting to explain the approximation methods, it's necessary to address some preliminaries – the basic mathematical paradigms that we'll use. As such, let's assume that the stationary, Gaussian, random process of interest is characterized in state-space as a typical linear system of the form

$$\begin{aligned} \mathbf{q}_{k+1} &= \mathbf{A}\mathbf{q}_k + \mathbf{B}n_k \\ x_k &= \mathbf{C}\mathbf{q}_k + v_k \end{aligned} \tag{3.6}$$

where $\mathbf{q}_k$ is the unobserved state of the process with measured output x_k. Apriori statistics for the input and measurement noise sequences, n_k and v_k, also need to be defined. Their covariances are

$$\begin{aligned} Q &\triangleq E[n_k n_k^T] = \sigma_n^2 \\ R &\triangleq E[v_k v_k^T] \end{aligned}$$

We also assume n_k and v_k are zero-mean Gaussian white noise sequences without loss of generality, such that $n_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$. At this point we should introduce the propagation of the *unconditional* covariance matrix.

$$\mathbf{P}_{k+1} = \mathbf{A}\mathbf{P}_k\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T \tag{3.7}$$

$$= \mathbf{A}\mathbf{P}_k\mathbf{A}^T + \mathbf{B}\sigma_n^2\mathbf{B}^T \tag{3.8}$$

where

$$\mathbf{P}_k \stackrel{\triangle}{=} E[\mathbf{q}_k \mathbf{q}_k^T]$$

There is also an algebraic equivalent to this propagation equation.

$$\mathbf{P}_{ss}^L = \mathbf{A}\mathbf{P}_{ss}^L\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T = \mathbf{A}\mathbf{P}_{ss}^L\mathbf{A}^T + \mathbf{B}\sigma_n^2\mathbf{B}^T$$

where $\mathbf{P}_{ss}^L \succ 0$ is the solution to this discrete algebraic Lyapunov equation. If we take the conditional expectation of Eqn. 3.6 with respect to the data, $\mathcal{D} = \{x_0, \ldots, x_k\}$, we obtain:

$$\hat{\mathbf{q}}_{k+1|k} = \mathbf{A}\hat{\mathbf{q}}_{k|k} \tag{3.9}$$

where

$$\hat{\mathbf{q}}_{k|k} \stackrel{\triangle}{=} E[\mathbf{q}_k | x_0, \ldots, x_k]$$

This represents the time update step of the recursive Kalman filtering equations. The measurement update step is given by

$$\hat{\mathbf{q}}_{k+1|k+1} = \hat{\mathbf{q}}_{k+1|k} + \mathbf{P}_{k+1|k}\mathbf{C}^T(\mathbf{C}\mathbf{P}_{k+1|k}\mathbf{C}^T + \mathbf{R})^{-1}(x_{k+1} - \mathbf{C}\hat{\mathbf{q}}_{k+1|k}) \tag{3.10}$$

We also need to define the *conditional* covariance propagation time and measurement updates in Eqns. 3.11-3.12, respectively.

$$\mathbf{P}_{k+1|k} = \mathbf{A}\mathbf{P}_{k|k}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T \tag{3.11}$$

$$\mathbf{P}_{k+1|k+1} = \mathbf{P}_{k+1|k} - \mathbf{P}_{k+1|k}\mathbf{C}^T(\mathbf{C}\mathbf{P}_{k+1|k}\mathbf{C}^T + \mathbf{R})^{-1}\mathbf{C}\mathbf{P}_{k+1|k} \tag{3.12}$$

where

$$\mathbf{P}_{k|k} \stackrel{\triangle}{=} E[(\mathbf{q}_k - \hat{\mathbf{q}}_{k|k})(\mathbf{q}_k - \hat{\mathbf{q}}_{k|k})^T | x_0, \ldots, x_k]$$

Combining the two equations, we get the following:

$$\mathbf{P}_{k+1|k} = \mathbf{A}\mathbf{P}_{k|k-1}\mathbf{A}^T - \mathbf{A}\mathbf{P}_{k|k-1}\mathbf{C}^T(\mathbf{C}\mathbf{P}_{k|k-1}\mathbf{C}^T + \mathbf{R})^{-1}\mathbf{C}\mathbf{P}_{k|k-1}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T \tag{3.13}$$

The stationary version of Eqn. 3.13 gives us the solution to the discrete algebraic Riccati equation, as follows:

$$\mathbf{P}_{ss}^R = \mathbf{A}\mathbf{P}_{ss}^R\mathbf{A}^T - \mathbf{A}\mathbf{P}_{ss}^R\mathbf{C}^T(\mathbf{C}\mathbf{P}_{ss}^R\mathbf{C}^T + \mathbf{R})^{-1}\mathbf{C}\mathbf{P}_{ss}^R\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T \tag{3.14}$$

However, we're interested in the updated aposteriori steady-state covariance matrix, which is the stationary version of Eqn. 3.12 given by:

$$\hat{\mathbf{P}}_{ss}^R = \mathbf{P}_{ss}^R - \mathbf{P}_{ss}^R\mathbf{C}^T(\mathbf{C}\mathbf{P}_{ss}^R\mathbf{C}^T + \mathbf{R})^{-1}\mathbf{C}\mathbf{P}_{ss}^R \tag{3.15}$$

In Eqn. 3.15, it is immediately apparent that $\hat{\mathbf{P}}_{ss}^R \prec \mathbf{P}_{ss}^R$ in the quadratic sense. There are some other important Kalman filter properties which give us more definiteness bounds.

Theorem 1 *Suppose (**A**,**C**) is observable, then $\boldsymbol{P}_{k+1|k}$ is bounded, i.e. $\exists \boldsymbol{M} : \boldsymbol{P}_{k+1|k} \prec \boldsymbol{M}$.*

Theorem 2 *Suppose (**A**,**C**) is observable, then $\boldsymbol{P}_{k+1|k} \rightarrow \boldsymbol{P}_{ss}^R,\ \forall\ \boldsymbol{P}_{0|-1}$.*

Theorem 3 *Suppose (**A**,**B**) is controllable, then the Discrete Algebraic Riccati Equation (DARE) has a unique solution, $\boldsymbol{P}_{ss}^R \succ 0$.*

Theorem 4 *Suppose (**A**,**B**) is controllable and (**A**,**C**) is observable, then $\exists$ a unique $\boldsymbol{P}_{ss}^R \succ 0$ which solves the DARE.*

Therefore, if (**A**,**B**) is controllable and (**A**,**C**) is observable, then the solution to the DARE will be positive definite, and $\mathbf{P}_{ss}^R \succ 0$. It may also be shown that $\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R \succ 0$, which will be needed for future analysis.

$\hat{\mathbf{P}}_{ss}^R$ can be used in place of $\mathbf{P}_{k|k}$, and $\mathbf{P}_{ss}^L$ can be used in place of $\mathbf{P}_k$, because that's what they will converge to anyway. Since our process is assumed to be stationary, and if we initialize

our Kalman filter to steady-state conditions, there will be no transient convergence period. Because we're primarily concerned with prediction for the main purposes of this chapter, we will need to compute variances and covariances of the form

$$V_{k+i|k} \triangleq Var(x_{k+i}|x_0, \ldots, x_k)$$

$$cov(x_{k+i}, x_{k+j}|x_0, \ldots, x_k)$$

where i, j are prediction window indices. It can be shown that $V_{k+i|k}$ and its covariance counterparts are functions of $\mathbf{P}_{k|k}$ and $\mathbf{P}_k$, and therefore $\hat{\mathbf{P}}_{ss}^R$, and $\mathbf{P}_{ss}^L$, respectively. As a result, they can be expressed as being independent of the time index k, although they cannot be expressed as being independent of the prediction window indices i, j. Similarly, we will need to compute the predicted process value $\hat{x}_{k+i|k}$.

$$\begin{aligned}
\hat{x}_{k+i|k} &\triangleq E[x_{k+i}|x_0, \ldots, x_k] && (3.16)\\
&= \mathbf{C}E[\mathbf{q}_{k+i}|x_0, \ldots, x_k] && (3.17)\\
&= \mathbf{C}\hat{\mathbf{q}}_{k+i|k} && (3.18)\\
&= \mathbf{C}(\mathbf{A}^i\hat{\mathbf{q}}_{k|k} + (\mathbf{I}_n - \mathbf{A}^i)(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) && (3.19)\\
\text{where } \mu_{T_B} &\triangleq E[x_k] && (3.20)
\end{aligned}$$

Obviously, $\hat{x}_{k+i|k}$ can be expressed as a function of $\hat{\mathbf{q}}_{k|k}$, but unlike $V_{k+i|k}$, it will fluctuate as new measurements are made. This is apparent due to Eqn. 3.10, which is directly dependent upon measurements x_{k+1}.

3.2.3 Approximations for the Exact Alarm Region

Two-Dimensional Alarm System Approximation

The first approximation for the exact alarm region given by $A_{exact} = \{\hat{\mathbb{X}} : P(x_k < L|\mathcal{D}) - P(x_k < L, x_{k+1} < L, \ldots, x_{k+m} < L|\mathcal{D}) \geq P_b\}$ is a single interval two-dimensional approximation.

In this case we only consider two time slices, $\{i = 0, i = m\}$, which span the entire time interval being considered. The process value is below the critical threshold at the very beginning of the interval, and above it at the very end. The exact null hypothesis captures at least one complaint or upcrossing, where the process itself is not restricted to being above or below the critical threshold at the end of the interval. Therefore, this approximation will miss half of the upcrossings/complaints caught by the exact condition. However, using it will greatly reduce computation time, and the answer can be achieved via numerical integration, without the use of simulation.

The approximation is shown in Eqns. 3.21-3.25. By using it, not only do we reduce the dimension of the alarm region from $m + 1$ to 2, shown in the first step via Eqn. 3.21, but a "rectangularized" two-dimensional approximation is used for further ease of computation. Recall that a two-dimensional alarm system is one that actually uses the idea of optimal alarm. The exact $2D$ alarm region is approximated with asymptotes, making the revised region of integration much easier to parameterize. The "rectangularization" is apparent in Fig. 3.1.

$$P(x_k < L, x_{k+m} > L | x_0, \ldots, x_k) \geq P_b \tag{3.21}$$

$$\Updownarrow \quad 2D \text{ approximation}$$

$$\hat{x}_{k|k} < L - \sqrt{V_{k|k}}\Phi^{-1}(P_b) \bigcap \hat{x}_{k+m|k} > L + \sqrt{V_{k+m|k}}\Phi^{-1}(P_b) \tag{3.22}$$

where

$$\Phi^{-1}(\cdot) = \text{Inverse cumulative normal standard distribution function} \tag{3.23}$$

$$V_{k+i|k} \triangleq Var(x_{k+i} | x_0, \ldots, x_k) \tag{3.24}$$

$$= \mathbf{C}(\mathbf{A}^i(\hat{\mathbf{P}}_{ss}^R - \mathbf{P}_{ss}^L)(\mathbf{A}^T)^i + \mathbf{P}_{ss}^L)\mathbf{C}^T + R \tag{3.25}$$

Notice that there are several two-dimensional alarm regions shown in Fig. 3.1, for values of P_b ranging from 0.1 to 0.9, in gradations of 0.1, all for $m = 5$. Each of the regions is convex, above

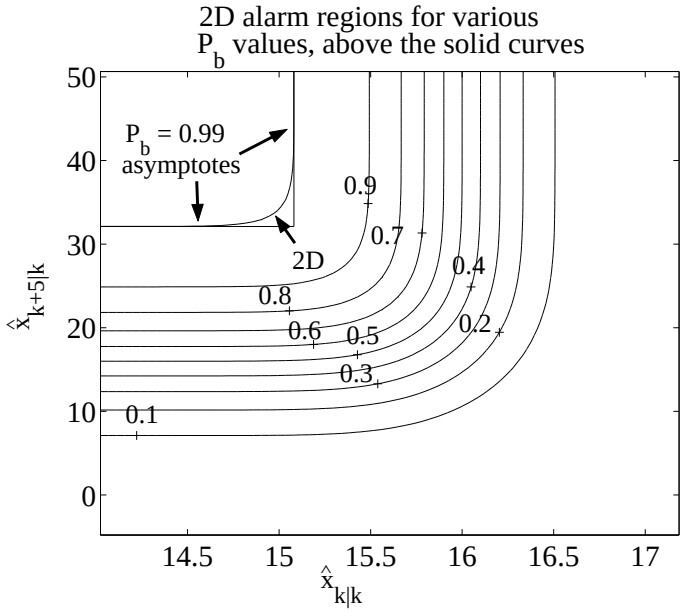

Figure 3.1: Approximation to the 2D Alarm Region for $P_b = 0.99$

and to the left of which is considered the alarm region. However, performing an integration over this two-dimensional alarm region requires storing the contour points along its border, and often consumes quite a bit more compute cycles than the respective "rectangularized" $2D$ alarm region. These contour plots can also be expressed approximately as a parametric function of $\hat{x}_{k|k}$ and $\hat{x}_{k+i|k}$. However, this approximation is best when the correlation between $\hat{x}_{k|k}$ and $\hat{x}_{k+i|k}$ nears 1. Also, using this function to perform an integration over the two-dimensional alarm region will still be quite a bit more computationally demanding than using the respective "rectangularized" $2D$ alarm region. For the specific case of $P_b = 0.99$, we can see the asymptotes in Fig. 3.1 that define the approximation, which bound the two-dimensional alarm region. Integration over this region is much easier, and the magnitude of the error introduced by this additional approximation will not be on par with the approximation of the "$m + 1$"-dimensional exact region with a reduced 2-dimensional region.

Multiple Sub-Interval Alarm System Approximation

To improve upon the approximation introduced in the previous section, and "catch" more of the cases missed by the single two-dimensional interval, we can split the interval into N_s disjoint two-dimensional subintervals, and construct alarm systems for each subinterval. Here we'd like to compute the relevant aggregate Type I/II error probabilities for the entire interval in question, by taking the *union* of the alarm systems corresponding to each subinterval. Obtaining the two-dimensional approximate alarm regions for each subinterval is easy, and is based upon the same logic in Eqns. 3.21-3.25. In order to determine the aggregate Type I/II error probabilities, we use Eqns. 3.30-3.33. But first the alarm sub-interval approximation definitions provided in Eqns. 3.26-3.29 must be made.

First, recall:

$$C_{exact} = \neg\{x_k < L, x_{k+1} < L, \ldots, x_{k+m} < L\} \setminus \{x_k > L\} \tag{3.26}$$

Alarm sub-interval approximation:

$$A_i = \{\hat{x}_{k+s_i|k} < \mathcal{X}_{s_i}, \hat{x}_{k+s_{i+1}|k} > \mathcal{Y}_{s_{i+1}}\} \tag{3.27}$$

$$\mathcal{X}_{s_i} = L - \sqrt{V_{k+s_i|k}}\Phi^{-1}(P_b) \tag{3.28}$$

$$\mathcal{Y}_{s_{i+1}} = L + \sqrt{V_{k+s_{i+1}|k}}\Phi^{-1}(P_b) \tag{3.29}$$

$$0 \leq s_i, s_{i+1} \leq m \quad \Leftrightarrow \quad \text{indices corresponding to endpoints of subinterval } i$$

Formulae for correct/false alarms and correct/missed detections can be developed, as shown in Eqns. 3.30-3.33.

$$\text{Correct Alarm: } P(C_{exact}|\bigcup_i A_i) = \frac{P(C_{exact}, \bigcup_i A_i)}{P(\bigcup_i A_i)} \tag{3.30}$$

$$\text{Correct Detection: } P(\bigcup_i A_i|C_{exact}) = \frac{P(C_{exact}, \bigcup_i A_i)}{P(C_{exact})} \tag{3.31}$$

$$\begin{aligned} \text{False Alarm: } P(C'_{exact}|\bigcup_i A_i) &= \frac{P(C'_{exact}, \bigcup_i A_i)}{P(\bigcup_i A_i)} \\ &= 1 - P(C_{exact}|\bigcup_i A_i) \end{aligned} \tag{3.32}$$

$$\begin{aligned} \text{Missed Detection: } P((\bigcup_i A_i)'|C_{exact}) &= \frac{P(C_{exact}, (\bigcup_i A_i)')}{P(C_{exact})} \\ &= 1 - P(\bigcup_i A_i|C_{exact}) \end{aligned} \tag{3.33}$$

In order to compute the deceivingly simple-looking formulae in Eqns. 3.30-3.33, we need to come up with more detailed equations for $P(\bigcup_i A_i)$, $P(C_{exact}, \bigcup_i A_i)$, and $P(C_{exact})$.

$$
\begin{aligned}
P(C_{exact}) &= P(x_k < L) - P(x_k < L, \ldots, x_{k+m} < L) && (3.34)\\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{L} e^{-\frac{1}{2}\frac{(x_k - \mathbf{C}\mu_\mathbf{q})^2}{\mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + \mathbf{R}}} dx_k - \underbrace{\int_{-\infty}^{L} \cdots \int_{-\infty}^{L}}_{m+1} \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x}\\
&= \Phi\left(\frac{L - \mathbf{C}\mu_\mathbf{q}}{\sqrt{\mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + \mathbf{R}}}\right) - \underbrace{\int_{-\infty}^{L} \cdots \int_{-\infty}^{L}}_{m+1} \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})\ d\mathbf{x}\\
\mu_\mathbf{x} &= \begin{bmatrix} \mathbf{C}\mu_\mathbf{q} \\ \vdots \\ \mathbf{C}(\mathbf{A}^i\mu_\mathbf{q} + (\mathbf{I}_n - \mathbf{A}^i)(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \\ \vdots \\ \mathbf{C}(\mathbf{A}^m\mu_\mathbf{q} + (\mathbf{I}_n - \mathbf{A}^m)(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mu_\mathbf{q} \\ \vdots \\ \mathbf{C}\mu_\mathbf{q} \\ \vdots \\ \mathbf{C}\mu_\mathbf{q} \end{bmatrix}\\
\Sigma_\mathbf{x} &= \begin{bmatrix} \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + \mathbf{R} & \cdots & \mathbf{C}\mathbf{P}_{ss}^{L}(\mathbf{A}^T)^i\mathbf{C}^T & \cdots & \mathbf{C}\mathbf{P}_{ss}^{L}(\mathbf{A}^T)^m\mathbf{C}^T \\ \vdots & \ddots & & & \vdots \\ \mathbf{C}\mathbf{A}^i\mathbf{P}_{ss}^{L}\mathbf{C}^T & \cdots & \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + \mathbf{R} & \cdots & \mathbf{X}_{ss} \\ \vdots & & & \ddots & \vdots \\ \mathbf{C}\mathbf{A}^m\mathbf{P}_{ss}^{L}\mathbf{C}^T & \cdots & \mathbf{X}_{ss}^T & \cdots & \mathbf{C}\mathbf{P}_{ss}^{L}\mathbf{C}^T + \mathbf{R} \end{bmatrix}\\
\mathbf{X}_{ss} &= \mathbf{C}\mathbf{L}_{ss}\mathbf{C}^T && (3.35)\\
\mathbf{L}_{ss} &= \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T(\mathbf{A}^T)^{m-i} \text{ where } i < m && (3.36)\\
\mathbf{x} &= \begin{bmatrix} x_k \\ \vdots \\ x_{k+m} \end{bmatrix} \in \mathbb{R}^{m+1} && (3.37)\\
\mu_\mathbf{q} &\triangleq E[\mathbf{q}_k] && (3.38)
\end{aligned}
$$

where $\mathcal{N}(\mathbf{x}; \mu, \Sigma)$ is the standard multivariate normal probability distribution of $\mathbf{x}$ with mean $\mu_\mathbf{x}$, and covariance matrix $\Sigma_\mathbf{x}$. Note that $\mathbf{X}_{ss}$ is the cross term of covariance matrix $\Sigma_\mathbf{x}$ corresponding to the i^{th} row and $(m+1)^{st}$ column. However, the general formula for cross terms is $\mathbf{C}\mathbf{L}_{ss}\mathbf{C}^T$, where

$i < j$ for the i^{th} row and the j^{th} column, and $\mathbf{L}_{ss} = \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T(\mathbf{A}^T)^{j-i}$.

Let $A_{approx} \triangleq \bigcup_{i=1}^{N_s} A_i$ denote the alarm condition for the sub-interval approximation, and a candidate region of integration. Note that the integrands of Eqns. 3.39 and 3.41 cannot easily be integrated over this region of integration. Since these integrals are not well-defined, we must use the inclusion/exclusion rule to develop more extensive formulae for $P(\bigcup_i A_i)$ and $P(C_{exact}, \bigcup_i A_i)$.

However, we know that often an alternative to using the inlcusion/exclusion rule, which has $\mathcal{O}(2^{N_s})$ terms, is to compute its complement instead, i.e. $P(\bigcup_i A_i) = 1 - P(\bigcap_i A_i')$. This is often a much less computationally intensive formula to work with, and less numerical integration error accumulates, when considering the number of terms involved. Recall $A_i = \{\hat{x}_{k+s_i|k} < \mathcal{X}_{s_i}, \hat{x}_{k+s_{i+1}|k} > \mathcal{Y}_{s_{i+1}}\}$. Let's redefine $A_i \triangleq X_i \bigcap Y_i$, where $X_i \triangleq \{\hat{x}_{k+s_i|k} < \mathcal{X}_{s_i}\}$, and $Y_i \triangleq \{\hat{x}_{k+s_{i+1}|k} > \mathcal{Y}_{s_{i+1}}\}$. We know from DeMorgan's theorem that $A_i' = (X_i \bigcap Y_i)' = X_i' \bigcup Y_i'$. Therefore, $\bigcap_i A_i' = \bigcap_i \{X_i' \bigcup Y_i'\}$. This has $\mathcal{O}(N_s)$ terms, which is less than $\mathcal{O}(2^{N_s})$ terms, but the probability of this logical expression cannot be computed directly. It will need to be manipulated further in order to feasibly compute the probability, as follows:

$$\begin{aligned}
\bigcap_{i=1}^{N_s} A_i' &= \bigcap_{i=1}^{N_s} X_i' \bigcup Y_i' \\
&= \bigcup_{i=1}^{2^{N_s}} M_i \bigcap N_i \\
\text{where } (M_i, N_i) &\in \{(X_i, Y_i), (X_i, Y_i'), (X_i', Y_i), (X_i', Y_i')\}
\end{aligned}$$

So, we now compute $1 - P(\bigcap_{i=1}^{N_s} A_i') = 1 - P(\bigcup_{i=1}^{2^{N_s}} M_i \bigcap N_i)$, which requires $\mathcal{O}(2^{2^{N_s}})$ terms by using the inclusion/exclusion rule. But this is much more than the $\mathcal{O}(2^{N_s})$ terms required by computing $P(\bigcup_i A_i)$ directly. Therefore, we will compute $P(\bigcup_i A_i)$ in lieu of $1 - P(\bigcap_i A_i')$, for multiple sub-interval probability computations. We continue, now with Eqns. 3.39 and 3.41, by using the inclusion/exclusion rule:

$$P(\bigcup_{i=1}^{N_s} A_i) = \int_{\Omega_{A_{approx}}} \mathcal{N}(\hat{\mathbb{X}}; \mu_{\hat{\mathbb{X}}}, \Sigma_{\hat{\mathbb{X}}})\ d\hat{\mathbb{X}} \tag{3.39}$$

$$= \sum_{i=1}^{N_s} P(A_i) - \sum_{1 \le i < j \le N_s} P(A_i, A_j) \tag{3.40}$$

$$+ \sum_{1 \le i < j < k \le N_s} P(A_i, A_j, A_k) - \ldots + (-1)^{N_s - 1} P(\bigcap_{i=1}^{N_s} A_i)$$

$$\text{Let } \mathbf{x} = \begin{bmatrix} \mathbb{X} \\ \hat{\mathbb{X}} \end{bmatrix}$$

$$\text{Let } \mathbf{x}_{sub} = \begin{bmatrix} x_k \\ \hat{\mathbb{X}} \end{bmatrix}$$

$$P(C_{exact}, \bigcup_{i=1}^{N_s} A_i) = \int_{-\infty}^{L} \int_{\Omega_{A_{approx}}} \mathcal{N}(\mathbf{x}_{sub}; \mu_{\mathbf{x}_{sub}}, \Sigma_{\mathbf{x}_{sub}})\ d\mathbf{x}_{sub} \tag{3.41}$$

$$- \underbrace{\int_{-\infty}^{L} \cdots \int_{-\infty}^{L}}_{m+1} \int_{\Omega_{A_{approx}}} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x}$$

$$= \sum_{i=1}^{N_s} P(C_{exact}, A_i) - \sum_{1 \le i < j \le N_s} P(C_{exact}, A_i, A_j) \tag{3.42}$$

$$+ \sum_{1 \le i < j < k \le N_s} P(C_{exact}, A_i, A_j, A_k) - \ldots + (-1)^{N_s - 1} P(C_{exact}, \bigcap_{i=1}^{N_s} A_i)$$

The region of integration, A_{approx}, has the following representation:

$$
\begin{aligned}
A_{approx} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{approx}} \subset \mathbb{R}^{N_s+1}\} \qquad (3.43)\\
&= \{\hat{\mathbb{X}} : \bigcup_{i=1}^{N_s} A_i\}\\
\text{Let } C_i &= \{x_{k+s_i} < L, x_{k+s_{i+1}} > L\}\\
\Rightarrow A_{approx} &= \{\hat{\mathbb{X}} : \bigcup_{i=1}^{N_s} \{P(C_i|\mathcal{D}) \geq P_b\}\}\\
&= \{\hat{\mathbb{X}} : \bigcup_{i=1}^{N_s} \{P(x_{k+s_i} < L, x_{k+s_{i+1}} > L|x_0, \ldots, x_k) \geq P_b\}\}\\
&\overset{\text{Subinterval Approximation}}{=} \{\hat{\mathbb{X}} : \bigcup_{i=1}^{N_s} \{\hat{x}_{k+s_i|k} < \mathcal{X}_{s_i}, \hat{x}_{k+s_{i+1}|k} > \mathcal{Y}_{s_{i+1}}\}\}\\
\text{Recall } \mathcal{X}_{s_i} &= L - \sqrt{V_{k+s_i|k}}\Phi^{-1}(P_b)\\
\text{and } \mathcal{Y}_{s_{i+1}} &= L + \sqrt{V_{k+s_{i+1}|k}}\Phi^{-1}(P_b)
\end{aligned}
$$

The algorithm used to form different combinations of A_i, $i \in \{1 \ldots m\}$ for use in Eqns. 3.40 and 3.42 is as follows:

1. Form the set $\mathcal{A} = \{A_1, A_2, \ldots, A_i, \ldots, A_m\}$.

2. Extract vacuous and singleton elements of its powerset, $2^{\mathcal{A}}$.

3. The remaining non-vacuous and non-singleton elements represent the pairs, triples, etc., necessary to compute Eqns. 3.40 and 3.42.

Examples of some of the probabilities needed to compute $P(\bigcup_i A_i)$ are as follows:

$$P(A_i) = P(\hat{x}_{k+s_i|k} < \mathcal{X}_{s_i}, \hat{x}_{k+s_{i+1}|k} > \mathcal{Y}_{s_{i+1}}) \tag{3.44}$$

$$= \int_{-\infty}^{\mathcal{X}_{s_i}} \int_{\mathcal{Y}_{s_{i+1}}}^{\infty} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} \tag{3.45}$$

$$\mu_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}(\mathbf{A}^{s_i}\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^{s_i})(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \\ \mathbf{C}(\mathbf{A}^{s_{i+1}}\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^{s_{i+1}})(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \tag{3.46}$$

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mathbf{A}^{s_i} \\ \mathbf{C}\mathbf{A}^{s_{i+1}} \end{bmatrix} (\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R}) \begin{bmatrix} \mathbf{C}\mathbf{A}^{s_i} \\ \mathbf{C}\mathbf{A}^{s_{i+1}} \end{bmatrix}^T \tag{3.47}$$

$$\mathbf{x} = \begin{bmatrix} \hat{x}_{k+s_i|k} \\ \hat{x}_{k+s_{i+1}|k} \end{bmatrix} \in \mathbb{R}^2 \tag{3.48}$$

For the last term in Eqn. 3.40, in which the maximum number of alarm sub-intervals are used ($N_s = m$), and are all adjacent, the formula is shown below:

$$
\begin{aligned}
P(A_1, A_2, \ldots, A_m) &= P(\hat{x}_{k+s_1|k} < \mathcal{X}_{s_1}, \hat{x}_{k+s_2|k} > \mathcal{Y}_{s_2}, \\
&\quad \hat{x}_{k+s_2|k} < \mathcal{X}_{s_2}, \hat{x}_{k+s_3|k} > \mathcal{Y}_{s_3}, \ldots, \\
&\quad \hat{x}_{k+s_m|k} < \mathcal{X}_{s_m}, \hat{x}_{k+s_{m+1}|k} > \mathcal{Y}_{s_{m+1}}) \\
&= \int_{-\infty}^{\mathcal{X}_{s_1}} \int_{\mathcal{Y}_{s_2}}^{\mathcal{X}_{s_2}} \int_{\mathcal{Y}_{s_3}}^{\mathcal{X}_{s_3}} \cdots \int_{\mathcal{Y}_{s_m}}^{\mathcal{X}_{s_m}} \int_{\mathcal{Y}_{s_{m+1}}}^{\infty} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \, d\mathbf{x}
\end{aligned} \tag{3.49}
$$

$$
\begin{aligned}
\mu_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C}(\mathbf{A}^{s_1}\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^{s_1})(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \\ \vdots \\ \mathbf{C}(\mathbf{A}^{s_{m+1}}\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^{s_{m+1}})(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \vdots \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \\
\Sigma_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C}\mathbf{A}^{s_1} \\ \vdots \\ \mathbf{C}\mathbf{A}^{s_{m+1}} \end{bmatrix} (\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R}) \begin{bmatrix} \mathbf{C}\mathbf{A}^{s_1} \\ \vdots \\ \mathbf{C}\mathbf{A}^{s_{m+1}} \end{bmatrix}^T \\
\mathbf{x} &= \begin{bmatrix} \hat{x}_{k+s_1|k} \\ \vdots \\ \hat{x}_{k+s_{m+1}|k} \end{bmatrix} \in \mathbb{R}^{|\mathcal{A}|+1} \equiv \mathbb{R}^{m+1}
\end{aligned} \tag{3.50}
$$

Or for the last term in Eqn. 3.40, in which less than the maximum number of sub-intervals are used ($N_s = j < m$ or simply $j < m$), yet are still all adjacent, the formula is shown below:

$$
\begin{aligned}
\text{Let } \mathcal{A}_{sub} &= \{A_1, A_2, \ldots, A_j\} \subset \mathcal{A} \\
P(A_1, \ldots, A_j) &= P(\hat{x}_{k+s_1|k} < \mathcal{X}_{s_1}, \hat{x}_{k+s_2|k} > \mathcal{Y}_{s_2}, \ldots \\
&\quad \hat{x}_{k+s_j|k} < \mathcal{X}_{s_j}, \hat{x}_{k+s_{j+1}|k} > \mathcal{Y}_{s_{j+1}}) \\
&= \int_{-\infty}^{\mathcal{X}_{s_1}} \int_{\mathcal{Y}_{s_2}}^{\mathcal{X}_{s_2}} \cdots \int_{\mathcal{Y}_{s_j}}^{\mathcal{X}_{s_j}} \int_{\mathcal{Y}_{s_{j+1}}}^{\infty} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}) \, d\mathbf{x} \\
\mu_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C}(\mathbf{A}^{s_1}\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^{s_1})(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \\ \vdots \\ \mathbf{C}(\mathbf{A}^{s_{j+1}}\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^{s_{j+1}})(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \vdots \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \\
\Sigma_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C}\mathbf{A}^{s_1} \\ \vdots \\ \mathbf{C}\mathbf{A}^{s_{j+1}} \end{bmatrix} (\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R}) \begin{bmatrix} \mathbf{C}\mathbf{A}^{s_1} \\ \vdots \\ \mathbf{C}\mathbf{A}^{s_{j+1}} \end{bmatrix}^T \\
\mathbf{x} &= \begin{bmatrix} \hat{x}_{k+s_1|k} \\ \vdots \\ \hat{x}_{k+s_{j+1}|k} \end{bmatrix} \in \mathbb{R}^{|\mathcal{A}_{sub}|+1} \equiv \mathbb{R}^{j+1}
\end{aligned} \tag{3.51, 3.52}
$$

Mathematical Considerations

There are several important things to notice in Eqns. 3.44-3.52. The first is the fact that a lower limit need not be less than an upper limit for any component of a definite integral, i.e. there is no restriction which yields $\mathcal{Y}_{s_i} < \mathcal{X}_{s_i}$. In fact, it is quite possible that $\mathcal{Y}_{s_i} \geq \mathcal{X}_{s_i}$. When this is the case, mathematically computing the integrals shown in Eqns. 3.44-3.52 may result in negative numbers. However, from basic probability theory we know that $P(\cdot) \geq 0$, and $P(\emptyset) = 0$. Let's apply these ideas to obtain accurate answers for Eqns. 3.44-3.52. We know that $\forall i : \mathcal{Y}_{s_i} < \mathcal{X}_{s_i}$, $P(\cdot) > 0$, but for the specific case when $\exists i : \mathcal{Y}_{s_i} \geq \mathcal{X}_{s_i}$, the following is true:

$$\begin{aligned} P(\ldots, \hat{x}_{k+s_i|k} > \mathcal{Y}_{s_i}, \hat{x}_{k+s_i|k} < \mathcal{X}_{s_i}, \ldots) &= P(\ldots, \mathcal{Y}_{s_i} < \hat{x}_{k+s_i|k} < \mathcal{X}_{s_i}, \ldots) \\ &= P(\emptyset) \\ &= 0 \end{aligned}$$

Recall $|\mathcal{A}| = m$ and $|\mathcal{A}_{sub}| = j$ represent the number of sub-intervals in each respective set. The dimension of the states in Eqns. 3.50, 3.52 are $m+1$ and $j+1$, which are 1 greater than the dimensions of $|\mathcal{A}| = m$, and $|\mathcal{A}_{sub}| = j$, respectively. This is due to the fact that all sub-intervals in question are adjacent, and there are 2 endpoints for each sub-interval. It is important to bookkeep the dimensions of these sets to ensure that all covariance matrices are positive (semi-)definite. Recall that ($\mathbf{A}$,$\mathbf{B}$) controllable and ($\mathbf{A}$,$\mathbf{C}$) observable were necessary and sufficient conditions for the solution to the DARE to be positive definite. Therefore, given ($\mathbf{A}$,$\mathbf{B}$) controllable and ($\mathbf{A}$,$\mathbf{C}$) observable, $\mathbf{P}_{ss}^{R} \succ 0$. It may also be shown from this that $\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R} \succ 0$.

In general, from Kalman filter theory, we know that measurements decrease uncertainty. The "norm," or amplifying power of the covariance matrices is related to this uncertainty, because they represent prediction/estimation error. Therefore, we can set up a chain of corresponding definiteness relations, because definiteness can be mathematically related to the norm. The covariance which is the "least positive definite" (or "most negative definite") in some sense is the one that is the least "uncertain", and also the one that conditions on the greatest number of measurements.

The steady-state aposteriori covariance matrix $\hat{\mathbf{P}}_{ss}^{R}$ is based upon measurements up to time $k+1$, as in Eqn. 3.12. The steady-state apriori covariance matrix $\mathbf{P}_{ss}^{R}$ is based upon measurements up to time k, as in Eqn. 3.11. The steady-state unconditional covariance matrix $\mathbf{P}_{ss}^{L}$ is based upon no measurements at all, as in Eqn. 3.7. As a result, the covariance matrix with the "most information" is at the bottom of the chain of negative definiteness relations. This is due to the fact that its gain will be the least in norm, since it has the least uncertainty (lowest estimation error) . This would be the steady-state aposteriori covariance matrix, $\hat{\mathbf{P}}_{ss}^{R}$, since it conditions all the way out to $k+1$.

The next would be the steady-state apriori covariance matrix, $\mathbf{P}_{ss}^{R}$, since it only conditions out to k. The final covariance matrix in the chain of negative definiteness relations would be $\mathbf{P}_{ss}^{L}$, since it is based upon no measurements at all. To summarize we get the following: $0 \prec \hat{\mathbf{P}}_{ss}^{R} \prec \mathbf{P}_{ss}^{R} \prec \mathbf{P}_{ss}^{L}$. Mathematically, it follows that $\mathbf{P}_{ss}^{L} \succ \mathbf{P}_{ss}^{R} \succ \hat{\mathbf{P}}_{ss}^{R} \succ 0$, and that $\mathbf{P}_{ss}^{L} - \mathbf{P}_{ss}^{R} \succ 0$, $\mathbf{P}_{ss}^{R} - \hat{\mathbf{P}}_{ss}^{R} \succ 0$, and $\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R} \succ 0$. These are standard results from linear algebra, and the latter, $\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R} \succ 0$ is now needed.

All of the covariance matrices in the formulae above have the form $\mathbf{L}^T\mathbf{M}\mathbf{L}$. We now know that $\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R} \succ 0$. Therefore, a standard result from linear algebra tells us that $\mathbf{M} = \mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R} \succ 0 \Leftrightarrow \mathbf{L}^T\mathbf{M}\mathbf{L} \succeq 0$, or in the most restrictive case of $P(A_1, A_2, \ldots, A_m)$, $\mathbf{M} = \mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R} \succ 0 \Leftrightarrow \mathbf{L}^T\mathbf{M}\mathbf{L} \succ 0$ iff $rank(\mathbf{L}) = |\mathcal{A}| = m + 1$. The ($\mathbf{A}$,$\mathbf{C}$) observability condition required to show that $\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R} \succ 0$ can also be written as a rank condition on the observability matrix, $rank(\begin{bmatrix} \mathbf{C} & \mathbf{CA} & \ldots & \mathbf{CA}^{n-1} \end{bmatrix}^T) = n$, where $n = rank(\mathbf{A})$. Notice that the "$\mathbf{L}^T$" portion of the covariance matrix in Eqn. 3.50 resembles the observability matrix. This is the case in which the maximum number of alarm sub-intervals are used ($N_s = m$), and are all adjacent. Hence all $\{s_1 \ldots s_{m+1}\}$ sub-interval indices are actually equivalent to $\{0 \ldots m\}$, and instead of $\mathbf{L} = \begin{bmatrix} \mathbf{CA}^{s_1} & \mathbf{CA}^{s_2} & \ldots & \mathbf{CA}^{s_{m+1}} \end{bmatrix}$, we have $\mathbf{L} = \begin{bmatrix} \mathbf{C} & \mathbf{CA} & \ldots & \mathbf{CA}^{m} \end{bmatrix}$.

To obtain a positive definite covariance matrix, we require $rank(\mathbf{L}) = |\mathcal{A}| = m+1$ because the most restrictive case of $P(A_1, A_2, \ldots, A_m)$ always needs to be computed. It is the last term in Eqn. 3.40, and needed for accurate computation of $P(\bigcup_{i=1}^{m} A_i)$ when being computed for the maximum number of sub-intervals. The only way that this is possible is if $rank(\mathbf{A}) = n = rank(\mathbf{L}) = m + 1$. This means that the order of the linear system in question must have $m + 1$ states, or the prediction window must be limited to $n-1$ time slices. Therefore, we have a necessary and sufficient condition for accurate computation of $P(\bigcup_{i=1}^{m} A_i)$, assuming that all covariance matrices need to be positive definite. However, the positive semi-definite condition $\mathbf{L}^T(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})\mathbf{L} \succeq 0$ always holds, for any case from the least to the most restrictive. The multivariate normal probability integrals are computed by using a Monte Carlo technique developed by Genz [27] (i.e. as shown in Eqns. 3.49,

3.51). Monte Carlo techniques are based on sampling, therefore a certain pre-selected number of sample points are used to obtain reasonable accuracy for the integration result. Using this technique allows us to compute probabilities of a degenerate Gaussian random vector having a singular, positive semi-definite covariance matrix.

As might be expected, similar formulae can be derived for intermediate terms with non-adjacent sub-intervals in the inclusion/exclusion formula for $P(\bigcup_i A_i)$. Furthermore, formulae to compute the terms in the inclusion/exclusion formula for the joint probability $P(C_{exact}, \bigcup_i A_i)$ can be found as well. The formulae given thus far should be sufficient to extrapolate these equations from the systematic framework provided.

Multi-Dimensional Alarm System Approximation

The final method introduced in this chapter is meant to provide the best possible approximation to the exact alarm region, given by $A_{exact} = \{\hat{\mathbb{X}} : P(x_k < L|\mathcal{D}) - P(x_k < L, x_{k+1} < L, \ldots, x_{k+m} < L|\mathcal{D}) \geq P_b\}$. We know that the exact alarm region does not serve as a well-defined region of integration. Therefore, the *exact* alarm system metrics cannot feasibly be obtained for complicated multi-dimensional events by means of numerical integration. To illustrate this fact, Fig. 3.2 shows the exact alarm region when $m = 3$ and the multivariate Gaussian integrand is 4 dimensional. The alarm region is *above* the surface shown.

It is apparent from the figure that the surface of the exact alarm region boundary is quite complex and does not serve as a feasible, parameterizable integration region. However, an approximate volume can be defined as a tight bound on the exact region via the unions and/or intersections of planes. This approximation forms a semi-infinite hyper-rectangular region in multi-dimensional space. For operating complaints only, the region of integration, A_{approx}, has the following representation:

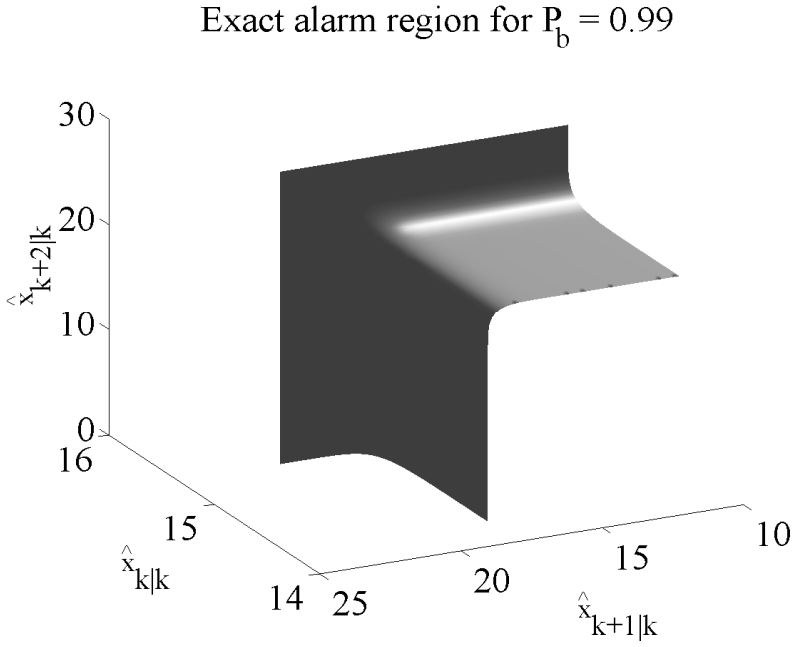

Figure 3.2: Exact Alarm Region for $m = 3$

$$
\begin{aligned}
A_{exact} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{exact}} \subset \mathbb{R}^{m+1}\} \\
&= \{\hat{\mathbb{X}} : P(x_k < L | x_0, \ldots, x_k) - \\
&\quad P(x_k < L, \ldots, x_{k+m} < L | x_0, \ldots, x_k) \geq P_b\} \\
&\Updownarrow \text{ Multi-dimensional approximation} \\
A_{approx} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{approx}} \subset \mathbb{R}^{m+1}\} \\
&= \{\hat{\mathbb{X}} : A_0 \bigcap \left[\bigcup_{i=1}^{m} A_i\right]\} \qquad (3.53) \\
\text{where } A_0 &\triangleq \hat{x}_{k|k} \leq \overbrace{L - \sqrt{V_{k|k}}\Phi^{-1}(P_b)}^{\mathcal{X}_0} \\
A_i &\triangleq \hat{x}_{k+i|k} \geq \underbrace{L + \sqrt{V_{k+i|k}}\Phi^{-1}(P_b)}_{\mathcal{Y}_i} \\
\forall i &\in 1, \ldots, m \\
\text{Or, } A_{approx} &= \{\hat{\mathbb{X}} : \hat{x}_{k|k} \leq L - \sqrt{V_{k|k}}\Phi^{-1}(P_b) \bigcap \left[\bigcup_{i=1}^{m} \hat{x}_{k+i|k} \geq L + \sqrt{V_{k+i|k}}\Phi^{-1}(P_b)\right]\}
\end{aligned}
$$

For both arrival and operating complaints, the region of integration, A_{approx}, has the following representation:

$$\begin{aligned}
A_{exact} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{exact}} \subset \mathbb{R}^{m+1}\} \\
&= \{\hat{\mathbb{X}} : 1 - P(x_{k+d} < L, \ldots, x_{k+d+m} < L | x_0, \ldots, x_k) \geq P_b\} \\
&\Updownarrow \text{ Multi-dimensional approximation} \\
A_{approx} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{approx}} \subset \mathbb{R}^{m+1}\} \\
&= \{\hat{\mathbb{X}} : \bigcup_{i=0}^{m} A_i\} \\
&= \{\hat{\mathbb{X}} : \bigcup_{i=0}^{m} \hat{x}_{k+d+i|k} \geq L + \sqrt{V_{k+d+i|k}}\Phi^{-1}(P_b)\}
\end{aligned}$$

Hence, Eqn. 3.53 denotes the approximate "operating complaints only" alarm condition for the multi-dimensional approximation, and a candidate region of integration. The formulae for correct/false alarms and correct/missed detections can be developed, as shown in Eqns. 3.54-3.57.

$$\text{Correct Alarm: } P(C_{exact}|A_{approx}) = \frac{P(C_{exact}, A_{approx})}{P(A_{approx})} \tag{3.54}$$

$$\text{Correct Detection: } P(A_{approx}|C_{exact}) = \frac{P(C_{exact}, A_{approx})}{P(C_{exact})} \tag{3.55}$$

$$\begin{aligned}
\text{False Alarm: } P(C'_{exact}|A_{approx}) &= \frac{P(C'_{exact}, A_{approx})}{P(A_{approx})} \qquad (3.56)\\
&= 1 - P(C_{exact}|A_{approx})
\end{aligned}$$

$$\begin{aligned}
\text{Missed Detection: } P((A_{approx})'|C_{exact}) &= \frac{P(C_{exact}, (A_{approx})')}{P(C_{exact})} \qquad (3.57)\\
&= 1 - P(A_{approx}|C_{exact})
\end{aligned}$$

In order to compute the formulae in Eqns. 3.54-3.57, we need to come up with more detailed equations for $P(A_{approx}) = P(A_0, \bigcup_{i=1}^{m} A_i)$ and $P(C_{exact}, A_{approx}) = P(C_{exact}, A_0, \bigcup_{i=1}^{m} A_i)$. We may use the inclusion/exclusion rule to develop extensive formulae for $P(A_{approx})$ and $P(C_{exact}, A_{approx})$, shown in Eqns. 3.61-3.65. However, recall that an alternative to using the inlcusion/exclusion rule,

which has $\mathcal{O}(2^m)$ terms, is to compute its complement instead, i.e. $P(\bigcup_i A_i) = 1 - P(\bigcap_i A_i')$. This method may turn out to be more useful than for sub-intervals, because of the way that the alarm events are defined for the multi-dimensional method. Recall the definitions are based upon single time slices as follows: $A_0 = \{\hat{x}_{k|k} \leq L - \sqrt{V_{k|k}}\Phi^{-1}(P_b)\}$, and $A_i = \{\hat{x}_{k+i|k} \geq L + \sqrt{V_{k+i|k}}\Phi^{-1}(P_b)\}$, $\forall i \in 1, \ldots, m$. So, we now may compute $P(A_0) - P(A_0, \bigcap_{i=1}^m A_i')$ in lieu of $P(A_0, \bigcup_{i=1}^m A_i)$, which only requires $\mathcal{O}(m)$ terms in lieu of the $\mathcal{O}(2^m)$ terms needed by the inclusion/exclusion rule. A more detailed formula for $P(A_0) - P(A_0, \bigcap_{i=1}^m A_i')$ is provided in Eqn. 3.58.

$$
\begin{aligned}
P(A_{approx}) &= P(A_0) - P(A_0, \bigcap_{i=1}^{m} A_i) \\
&= P(\hat{x}_{k|k} < \mathcal{X}_0) - P(\hat{x}_{k|k} < \mathcal{X}_0, \hat{x}_{k+1|k} < \mathcal{Y}_1, \ldots, \hat{x}_{k+m|k} < \mathcal{Y}_m) \qquad (3.58) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mathcal{X}_0} e^{-\frac{1}{2}\frac{(\hat{x}_{k|k} - \mathbf{C}\mu_{\mathbf{q}})^2}{\mathbf{C}(\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R)\mathbf{C}^T}} d\hat{x}_{k|k} - \int_{-\infty}^{\mathcal{X}_0}\int_{-\infty}^{\mathcal{Y}_1}\cdots\int_{-\infty}^{\mathcal{Y}_m} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\, d\mathbf{x} \\
&= \Phi\left(\frac{\mathcal{X}_0 - \mathbf{C}\mu_{\mathbf{q}}}{\sqrt{\mathbf{C}(\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R)\mathbf{C}^T}}\right) - \int_{-\infty}^{\mathcal{X}_0}\int_{-\infty}^{\mathcal{Y}_1}\cdots\int_{-\infty}^{\mathcal{Y}_m} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})
\end{aligned}
$$

$$
\mu_{\mathbf{x}} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mathbf{A}\mu_{\mathbf{q}} + \mathbf{B}\mu_{T_B} \\ \vdots \\ \mathbf{C}(\mathbf{A}^m\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^m)(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \\ \vdots \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix}
$$

$$
\Sigma_{\mathbf{x}} = \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^m \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^m \end{bmatrix}^T
$$

$$
\mathbf{x} = \begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+1|k} \\ \vdots \\ \hat{x}_{k+m|k} \end{bmatrix} \in \mathbb{R}^{m+1}
$$

Again, similar formulae may be derived for $P(C_{exact}, \bigcup_i A_i)$, and for the other cases of interest (both arrival and operating complaints, etc.). However, there may still be instances in which using the inclusion/exclusion rule provides greater accuracy for a smaller number of Monte Carlo integration sample points. This will be addressed further in the discussion section. Therefore we continue now with Eqns. 3.61 and 3.65, illustrating the formulae required for the inclusion/exclusion rule.

$$
\begin{aligned}
P(A_{approx}) &= P(A_0, \bigcup_{i=1}^{m} A_i) && (3.59)\\
&= P(A_0 \bigcap \left[\bigcup_{i=1}^{m} A_i\right]) && (3.60)\\
&= \int_{\Omega_{A_{approx}}} \mathcal{N}(\hat{\mathbb{X}}; \mu_{\hat{\mathbb{X}}}, \Sigma_{\hat{\mathbb{X}}})\ d\hat{\mathbb{X}} && (3.61)\\
&= \sum_{i=1}^{m} P(A_0, A_i) - \sum_{1\le i<j\le m} P(A_0, A_i, A_j) && (3.62)\\
&+ \sum_{1\le i<j<k\le m} P(A_0, A_i, A_j, A_k) - \ldots + (-1)^{m-1} P(\bigcap_{i=0}^{m} A_i)
\end{aligned}
$$

$$
\text{Let } \mathbf{x} = \begin{bmatrix} \mathbb{X} \\ \hat{\mathbb{X}} \end{bmatrix}
$$

$$
\text{Let } \mathbf{x}_{sub} = \begin{bmatrix} x_k \\ \hat{\mathbb{X}} \end{bmatrix}
$$

$$
\begin{aligned}
P(C_{exact}, A_{approx}) &= P(C_{exact}, A_0, \bigcup_{i=1}^{m} A_i) && (3.63)\\
&= \int_{-\infty}^{L} \int_{\Omega_{A_{approx}}} \mathcal{N}(\mathbf{x}_{sub}; \mu_{\mathbf{x}_{sub}}, \Sigma_{\mathbf{x}_{sub}})\ d\mathbf{x}_{sub} && (3.64)\\
&- \underbrace{\int_{-\infty}^{L} \cdots \int_{-\infty}^{L}}_{m+1} \int_{\Omega_{A_{approx}}} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x}\\
&= \sum_{i=1}^{m} P(C_{exact}, A_0, A_i) - \sum_{1\le i<j\le m} P(C_{exact}, A_0, A_i, A_j) && (3.65)\\
&+ \sum_{1\le i<j<k\le m} P(C_{exact}, A_0, A_i, A_j, A_k) - \ldots + (-1)^{m-1} P(C_{exact}, \bigcap_{i=0}^{m} A_i)
\end{aligned}
$$

Examples of some of the probabilities needed to compute $P(A_{approx}) = P(A_0, \bigcup_{i=1}^{m} A_i)$ are as follows:

$$
\begin{aligned}
P(A_0, A_i) &= P(\hat{x}_{k|k} < \mathcal{X}_0, \hat{x}_{k+i|k} > \mathcal{Y}_i) && (3.66)\\
&= \int_{-\infty}^{\mathcal{X}_0} \int_{\mathcal{Y}_i}^{\infty} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} && (3.67)\\
\mu_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}(\mathbf{A}^i\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^i)(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} && (3.68)\\
\Sigma_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A}^i \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A}^i \end{bmatrix}^T && (3.69)\\
\mathbf{x} &= \begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+i|k} \end{bmatrix} \in \mathbb{R}^2 && (3.70)
\end{aligned}
$$

For the second term in Eqn. 3.62, the formula is shown below:

$$
\begin{aligned}
P(A_0, A_i, A_j) &= P(\hat{x}_{k|k} < \mathcal{X}_0, \hat{x}_{k+i|k} > \mathcal{Y}_i, \hat{x}_{k+j|k} > \mathcal{Y}_j) && (3.71)\\
&= \int_{-\infty}^{\mathcal{X}_0} \int_{\mathcal{Y}_i}^{\infty} \int_{\mathcal{Y}_j}^{\infty} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\ d\mathbf{x} && (3.72)\\
\mu_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}(\mathbf{A}^i\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^i)(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \\ \mathbf{C}(\mathbf{A}^j\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^j)(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} && (3.73)\\
\Sigma_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A}^i \\ \mathbf{C}\mathbf{A}^j \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A}^i \\ \mathbf{C}\mathbf{A}^j \end{bmatrix}^T && (3.74)\\
\mathbf{x} &= \begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+i|k} \\ \hat{x}_{k+j|k} \end{bmatrix} \in \mathbb{R}^3 && (3.75)
\end{aligned}
$$

For the last term in Eqn. 3.62, the formula is shown below:

$$
\begin{aligned}
P(\bigcap_{i=0}^{m} A_i) &= P(\hat{x}_{k|k} < \mathcal{X}_0, \hat{x}_{k+1|k} > \mathcal{Y}_1, \ldots, \hat{x}_{k+m|k} > \mathcal{Y}_m) && (3.76)\\
&= \int_{-\infty}^{\mathcal{X}_0} \int_{\mathcal{Y}_1}^{\infty} \cdots \int_{\mathcal{Y}_m}^{\infty} \mathcal{N}(\mathbf{x}; \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})\, d\mathbf{x} \\
\mu_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{CA}\mu_{\mathbf{q}} + \mathbf{B}\mu_{T_B} \\ \vdots \\ \mathbf{C}(\mathbf{A}^m\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^m)(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \\ \vdots \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} && (3.77)\\
\Sigma_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^m \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^m \end{bmatrix}^T && (3.78)\\
\mathbf{x} &= \begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+1|k} \\ \vdots \\ \hat{x}_{k+m|k} \end{bmatrix} \in \mathbb{R}^{m+1} && (3.79)
\end{aligned}
$$

As before with the sub-interval method, recall that $\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R \succ 0$. Therefore, $\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R \succ 0 \Leftrightarrow \mathbf{L}^T\mathbf{M}\mathbf{L} \succeq 0$, where $\mathbf{L} = \begin{bmatrix} \mathbf{C} & \mathbf{CA} & \cdots & \mathbf{CA}^m \end{bmatrix}$ in the most restrictive case. Again, formulae to compute the terms in the inclusion/exclusion formula for the joint probability $P(C_{exact}, A_0, \bigcup_{i=1}^m A_i)$ can be found by using the systematic framework provided.

3.3 Example Details

3.3.1 Examples 1 and 2

The first pair of examples to test the theory presented in the previous section is Svensson's 2^{nd}-order model, with a fixed critical threshold. Recall the basic state-space model from Eqn. 3.6:

$$\begin{aligned} \mathbf{q}_{k+1} &= \mathbf{A}\mathbf{q}_k + \mathbf{B}n_k \\ x_k &= \mathbf{C}\mathbf{q}_k + v_k \end{aligned}$$

where $\mathbf{q}_k$ is the unobserved state of the process with measured output x_k. The process has a discrete time interval of $\tau_s = 1$ sec in between time steps. Here,

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -0.9 & 1.8 \end{bmatrix} \tag{3.80}$$

$$\mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{3.81}$$

$$\mathbf{C} = \begin{bmatrix} 0.5 & 1 \end{bmatrix} \tag{3.82}$$

Apriori statistics for the input and measurement noise sequences, n_k and v_k, also need to be quantified. For this example, their covariances are

$$\begin{aligned} Q &\triangleq E[n_k n_k^T] = \sigma_n^2 = 1 \\ R &\triangleq E[v_k v_k^T] = 0.08 \end{aligned}$$

Therefore, n_k and v_k are zero-mean Gaussian white noise sequences with variances as shown above, such that $n_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$. The level-crossing problem is formulated by choosing the critical, fixed threshold, $L = 16$. A sample realization of the process and the fixed threshold is shown in Fig. 3.3.

Our problem is to design an optimal alarm system that predicts at least one operating complaint, as defined previously ($P(x_k < L|x_0, \ldots, x_k) - P(x_k < L, x_{k+1} < L, \ldots, x_{k+m} < L|x_0, \ldots, x_k) \geq P_b$), or at least one arrival and operating complaint, also defined previously (1 −

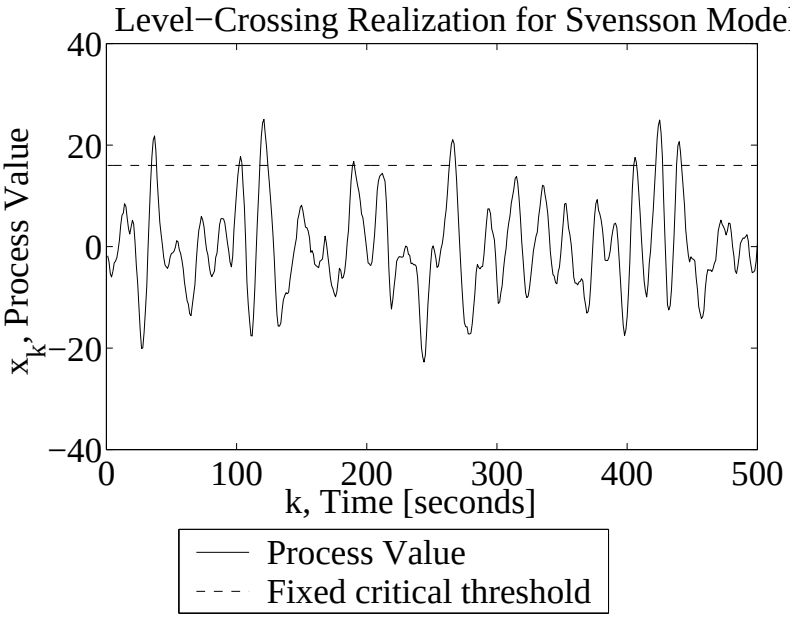

Figure 3.3: Process Realization and Level-Crossing Threshold for Examples 1, 2

$P(x_{k+d} < L, x_{k+d+1} < L, \ldots, x_{k+d+m} < L | x_0, \ldots, x_k) \geq P_b)$. Let's define these two cases as Example 1 and Example 2, respectively. For Example 1, we will use a 5-step ahead prediction window, or $m = 5$. For Example 2, we will use the same $m = 5$ step ahead prediction window, and $d = 2$ as the number of steps prior to the start of the beginning of the day.

3.3.2 Examples 3 and 4

The second pair of examples is based upon Federspiel's 2^{nd}-order model, which can be used for complaint prediction in thermal comfort applications. Recall that this model involves a random threshold, and the dynamics are quite different than those of Examples 1 or 2. In reality, there are two stochastic critical levels, one for hot complaints, and another for cold. These levels represent the temperature at which a group of occupants in a zone would complain if too hot or cold. However, only the hot complaint level is to be used as an example in this chapter to prevent redundancy. It is assumed that distinct optimal alarm systems for the two processes can be designed independently.

Therefore the hot complaint level was arbitrarily chosen as the example of interest. The state-space systems for both hot and cold critical levels, as well as the building temperature process of interest can easily be parameterized. We first define the two state-space equations for the complaint levels in continuous time:

$$\begin{aligned} \dot{\mathbf{z}}(t) &= \mathbf{A}\mathbf{z}(t) + \mathbf{B}n(t) \\ y(t) &= \mathbf{C}\mathbf{z}(t) \end{aligned}$$

The building temperature process is also defined first in continuous time:

$$\begin{aligned} \dot{\mathbf{q}}(t) &= \mathbf{A}\mathbf{q}(t) + \mathbf{B}u(t) + \mathbf{B}_w w(t) \\ x(t) &= \mathbf{C}\mathbf{q}(t) + v(t) \end{aligned}$$

$$\text{where } \mathbf{A} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -2\zeta\omega \end{bmatrix} \tag{3.83}$$

$$\mathbf{B} = \begin{bmatrix} 0 \\ \omega^2 \end{bmatrix}, \mathbf{B}_w = \begin{bmatrix} 0 \\ \omega^2 \end{bmatrix} \tag{3.84}$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{3.85}$$

Note that we use the standard second order canonical form to represent the state-space model for all three of the processes. It is possible to derive a formula, $\omega = \frac{\sigma_{\dot{T}}}{\sigma_T}$ by making the connection between the statistics of the process output, $y(t)$ and its derivative, $\dot{y}(t)$, as shown in [21, 22, 46]. Here, σ stands for standard deviation, and T, $\dot{T}$ stand for the output of any one of the three processes, and its derivative, respectively. Therefore, $T_B = x(t)$ for the building temperature process, or $T = y(t)$ for either of the complaint level processes. Typically, we will use the subscript B for the building temperature process, and the subscript H or L, for the hot and cold complaint

level processes, respectively. However, often we will leave this variable unsubscripted when there is no need to make a distinction among any of the processes, or when we are referring only to the hot complaint process.

All of these systems must be discretized for the discrete time analysis in the previous section to apply. This is done by performing a zero-order hold on all of the above systems, again as shown in [46]. As such, a sampling interval, τ_s, must also be chosen. As a rule of thumb, we choose a discrete time sampling interval based upon it being a fraction of the shortest time constant of the dynamics of all relevant processes. This will also be a useful fact when considering "whiteness" of the input and measurement noise. Specifically, $\tau_s = \frac{\min(\tau_h, \tau_c, \tau_b)}{5} \approx 18$ min in between time steps, where $\tau_h = \frac{1}{\omega_h}$, $\tau_c = \frac{1}{\omega_c}$, $\tau_b = \frac{1}{\omega_b}$ are the time constants of the hot, cold, and building processes, respectively. These constants are provided later, in Table 3.1. We round τ_s to 20 min, since its even multiple is 60 min, or an hour, which is a much more convenient unit of time to deal with than multiples of 18 min.

Again, apriori statistics for the zero-mean input and measurement noise processes, $w(t)$ and $v(t)$, can be quantified. $n(t)$ is a non-zero mean white noise process. All are scalar processes, such that $w(t), v(t), n(t) \in \mathbb{R}$. Note that $u(t) \in \mathbb{R}$ is a scalar fixed control input, which in our application acts as the thermostat setpoint. For this example, the variances of $n(t)$, $w(t)$, and $v(t)$ are as follows:

$$\begin{aligned} Q_w &\triangleq E[w(t)w(t)^T] \\ Q_n &\triangleq E[n(t)n(t)^T] \\ R &\triangleq E[v(t)v(t)^T] \end{aligned}$$

Therefore, with a Gaussian assumption on the input noise, we have the following: $w(t) \sim \mathcal{N}(0, Q_w)$, $v(t) \sim \mathcal{N}(0, R)$, and $n(t) \sim \mathcal{N}(\mu_T, Q_n)$. Note that R is only applicable to the building

temperature process, as it is the measured process, and $n(t)$ is only applicable to the complaint process, where there is no defined "control input." Since we assume that modelling the complaint levels requires no direct control input term, the mean of the input noise driving these processes is taken as the mean of the output. Also, no measurement noise needs to be modelled because there are no measurements of these unobserved processes. Furthermore, it is interesting to note that the complaint processes can be modelled as being statistically independent of the building temperature process for reasons as stated in Chap. 2 and Federspiel [19].

For the building process, typically the measurements come from a DDC (Direct Digital Control) system associated with more sophisticated commercial building systems. Otherwise, these types of measurements often come from micro-dataloggers that record the temperature for a preset period of time. Usually there is some sensor lag and/or noise associated with these measurements. It has been shown that the lag can be modelled simply as a first order transfer function [47], or dealt with otherwise [21]. In the latter case, the formula for the standard deviation of the rate of change of the building temperature is readjusted by using the formula $\sigma_{\dot{T}_B} = \frac{\sigma_{T_B}}{\max(\frac{1}{\omega_B} - 0.4, 0.4)}$, to account for the lag. However, here we do not model the sensor lag, and only model the measurement noise, $v(t)$, as a zero-mean white noise process with variance R.

Table 3.1 provides the relevant parameters and respective values for all three processes in continuous time. With the exception of the last two columns, these parameters come from Federspiel's work [22]. The fourth columns' (ω) values were derived from the first three using documented methods [21, 22, 46]. The fifth columns' (ζ) values were selected heuristically for simulation/analysis ease [46, 47].

	μ_T	σ_T	$\sigma_{\dot{T}}$	ω	ζ
Hot	91^oF	5.06^oF	$1.14\frac{^oF}{hr}$	$0.6645\frac{rad}{hr}$	1
Cold	50.43^oF	6.14^oF	$4.08\frac{^oF}{hr}$	$0.2253\frac{rad}{hr}$	1
Bldg	74^oF	3.57^oF	$0.91\frac{^oF}{hr}$*	$0.2682\frac{rad}{hr}$	1

*Typically requires sensor lag adjustment

Table 3.1: Table of Building and Complaint Temperature Statistics

Recall that the discretized systems must be computed by using the sampling interval, τ_s. The measurement noise was found by computing the pooled variance of the recordings of 88 independent dataloggers during a single-point ice bath calibration over the course of approximately 17.78 hours. The measurements have a correlation time of τ_d = 5-minute increments, and our pooled variance is $R_{d_{raw}} = 0.047519$. The approximate power spectral density can be computed by using the formula $R_{psd} \approx 2R_{d_{raw}}\tau_d$, [55]. In order to find the discretized measurement noise, we must then scale by the sampling interval, τ_s, giving us $R_d = \frac{R_{psd}}{\tau_s} \approx \frac{2R_{d_{raw}}\tau_d}{\tau_s}$, [55]. However, in order to use this approximation, we assume the noise is white compared to the sample period, i.e. the sample period is small with respect to system time constant(s). Therefore it must hold that $\tau_d \ll \tau_s \ll min(\tau_h, \tau_b, \tau_c)$. In this case, with τ_d = 5-min, τ_s = 20-minute, and $min(\tau_h, \tau_b, \tau_c) = 90$ min, the approximation is marginal.

We must pay close attention to the relationship among these sampling intervals when modelling measurement noise. The same requirements of $\tau_d \ll \tau_s \ll min(\tau_h, \tau_b, \tau_c)$ are not necessary for our input noise discretization procedure, because an alternative method is used, shown in Eqns. 3.93-3.105 for all relevant processes. For the assumption of whiteness of measurement noise, if there is ever a case in which τ_d is on the same order of magnitude, or larger than τ_s, then the measurement noise effects must be modelled as dynamics within the state of the system itself, and estimated along with the other states using Kalman filter techniques. It may also be possible to ignore this and treat the measurement noise as white, however the estimates will no longer be optimal [55]. Otherwise, we can ignore the noise altogether in which case it may be considered as "unmodelled dynamics." In our case, however, with τ_d = 5-min, τ_s = 20-minute, and $min(\tau_h, \tau_b, \tau_c) = 90$ min, the measurement noise can still (marginally) be modelled as white.

The only reason why we are concerned with computing the measurement noise now is because we need to use it to perform Kalman filtering and prediction. This is a prerequisite to implementing an optimal alarm system. The formulae relating the remaining discretized, discrete-time parameters to their continuous-time counterparts are shown in Eqns. 3.86-3.109. The state-

space parameters are discretized by performing a zero-order hold of all three processes. The input noise discretization results are shown for the building temperature process only, using a documented procedure [6, 35, 25, 55]. The input noise variances for the **complaint** level processes were not discretized, rather they were found by using discrete-time Lyapunov analysis of the discretized state-space system (shown in part below). A documented method [46] shows details on how to derive the discrete-time input noise based on a continuous and discrete time Lyapunov analysis of the statistics for these processes.

Discretization of state-space equations for all processes:

$$\mathbf{z}_{k+1} = \mathbf{A}_d \mathbf{z}_k + \mathbf{B}_d n_k \tag{3.86}$$

$$y_k = \mathbf{C}_d \mathbf{z}_k \tag{3.87}$$

$$\mathbf{q}_{k+1} = \mathbf{A}_d \mathbf{q}_k + \mathbf{B}_d u_k + w_k \tag{3.88}$$

$$x_k = \mathbf{C}_d \mathbf{q}_k + v_k \tag{3.89}$$

where

$$\mathbf{A}_d = e^{\mathbf{A}\tau_s} \tag{3.90}$$

$$\mathbf{B}_d = (e^{\mathbf{A}\tau_s} - \mathbf{I})\mathbf{A}^{-1}\mathbf{B} \tag{3.91}$$

$$\mathbf{C}_d = \mathbf{C} \tag{3.92}$$

Discretization of input and measurement noise parameters:

Complaint Process Input Noise:

$$
\begin{aligned}
n_k &\sim \mathcal{N}(\mu_T, Q_{n_d}) && (3.93)\\
Q_{n_d} &= E[n_k n_k^T] && (3.94)\\
&= 1364.3^o F \text{ (for hot complaint level)} && (3.95)\\
&= 2008.9^o F \text{ (for cold complaint level)} && (3.96)\\
&= \frac{\sigma_{T_B}^2}{\mathbf{C}_d \tilde{\mathbf{P}}_{ss}^L \mathbf{C}_d^T} && (3.97)\\
\text{where } \tilde{\mathbf{P}}_{ss}^L &\triangleq \frac{\mathbf{P}_{ss}^L}{Q_{n_d}} && (3.98)\\
&= \mathbf{A}_d \tilde{\mathbf{P}}_{ss}^L \mathbf{A}_d^T + \mathbf{B}_{w_d} \mathbf{B}_{w_d}^T && (3.99)
\end{aligned}
$$

Building Temperature Process Input Noise:

$$
\begin{aligned}
w_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{w_d}) && (3.100)\\
\mathbf{Q}_{w_d} &= \int_0^{\tau_s} e^{\mathbf{A}\lambda} \mathbf{B}_w Q_w \mathbf{B}_w^T e^{\mathbf{A}^T \lambda} d\lambda && (3.101)\\
&= \begin{bmatrix} 0.0091816 & 1.0994 \times 10^{-5} \\ 1.0994 \times 10^{-5} & 1.8369 \times 10^{-8} \end{bmatrix} && (3.102)\\
\text{where } Q_w &= \frac{\sigma_{T_B}^2}{\mathbf{C}\tilde{\mathbf{P}}_c \mathbf{C}^T} \approx 7.2 \times 10^5 \ ^o F && (3.103)\\
\text{and } \tilde{\mathbf{P}}_c &\triangleq \frac{\mathbf{P}_c}{Q_w} && (3.104)\\
\mathbf{A}\tilde{\mathbf{P}}_c + \tilde{\mathbf{P}}_c \mathbf{A}^T &= \mathbf{B}_w \mathbf{B}_w^T && (3.105)
\end{aligned}
$$

Building Temperature Process Measurement Noise:

$$
\begin{aligned}
v_k &\sim \mathcal{N}(0, R_d) && (3.106)\\
R_d &= E[v_k v_k^T] && (3.107)\\
&\approx \frac{2 R_{d_{raw}} \tau_d}{\tau_s} && (3.108)\\
&= 0.023759^o F && (3.109)
\end{aligned}
$$

The level-crossing problem can be viewed using two distinct paradigms. One way is to consider the measured process as crossing a random threshold. Another way to consider it is to take the difference between the two processes, and characterize its crossing interactions as fixed zero-level upcrossings of the level $L = 0$. For this to happen, the process must be defined as $z(t) \stackrel{\triangle}{=} x(t) - y(t)$, or $z_k \stackrel{\triangle}{=} x_k - y_k$ when considering the interaction between the hot complaint process and the building temperature process. A sample realization of all three processes interacting is shown in Fig. 3.4. In this figure, we present level crossings as defined by the first paradigm. Fig. 3.5 shows zero-level

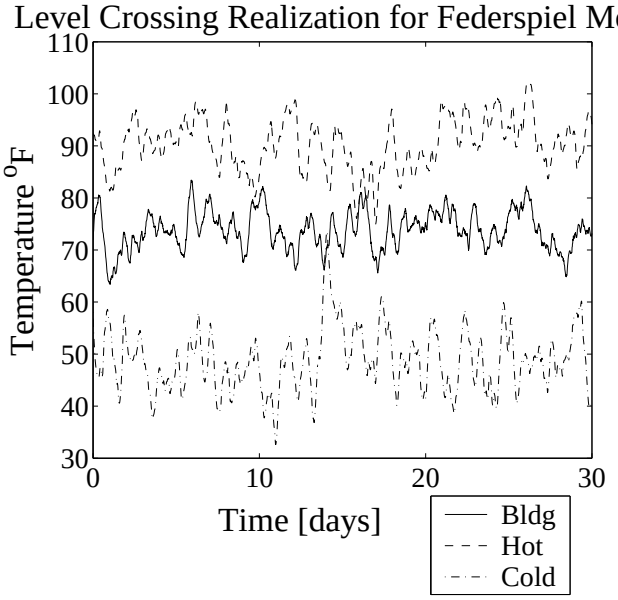

Figure 3.4: Building and Complaint Process Realizations with Level-Crossings

crossings of the difference between the hot complaint process and the building temperature process, which looks very similar to Fig. 3.3.

Again, our problem is to design an optimal alarm system that predicts at least one operating complaint, $P(z_k < 0|x_0, \ldots, x_k) - P(z_k < 0, z_{k+1} < 0, \ldots, z_{k+m} < 0|x_0, \ldots, x_k) \geq P_b)$, or at least one arrival and operating complaint, $1 - P(z_{k+d} < 0, z_{k+d+1} < 0, \ldots, z_{k+d+m} < 0|x_0, \ldots, x_k) \geq P_b$. Let's define these two cases as Example 3 and Example 4, respectively. For Example 3, we will use a

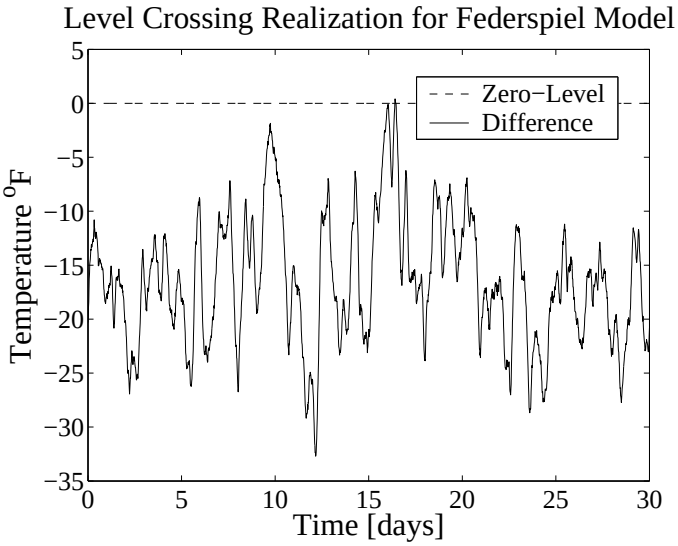

Figure 3.5: Building and Hot Complaint Process Realizations with Zero-Level Crossings

12-step ahead prediction window, or $m = 12$. This corresponds to a 4-hour period in which operating complaints can occur either during the morning or afternoon period of the day. For Example 4, we will use the same $m = 12$ step ahead prediction window, and $d = 3$ as the number of steps prior to the start of the day. In this 1-hour period prior to the beginning of the day, we want to predict both arrival complaints and operating complaints in the ensuing 4-hour period.

3.4 Results for Examples 1 and 2

3.4.1 Example 1: Svensson Model, Operating Complaints Only

Recall that complicated multi-dimensional events of the type proposed require simulations, and "counting" to obtain an estimate of the *exact* Type I/II error probabilities. The main point of the theory section was to provide viable alternative approximations to this "counting" method in order to reduce computation time. Although these approximations still require Monte Carlo simulations to compute numerical integrations of complex integrals, doing so is much less computationally

expensive than simulating and counting the number of correct/false alarms and correct/missed detections. Therefore, we shall simulate the probabilities using the exact method, and compute all of the approximations to the optimal alarm system for comparison. Our main goal is to determine which approximation is the most suitable by trading off accuracy with computational attractiveness. Recall that the optimal alarm systems of interest are the 2-dimensional alarm system approximation, the multiple sub-interval approximation (for all sub-intervals), and the multi-dimensional approximation. Clearly, there will be error involved with making any of these approximations. This error can be quantified by taking the 1-norm difference between the exact and approximate Type I/II errors across all border probabilities, P_b.

For this example, our problem is to design an optimal alarm system that predicts at least one operating complaint, $(P(x_k < L|x_0, \ldots, x_k) - P(x_k < L, x_{k+1} < L, \ldots, x_{k+m} < L|x_0, \ldots, x_k) \geq P_b$. Recall that we are using a 5-step ahead prediction window, or $m = 5$. An example of exact vs. approximate Type I/II errors for the 2-dimensional alarm system approximation as a function of border probabilities is shown in Fig. 3.6.

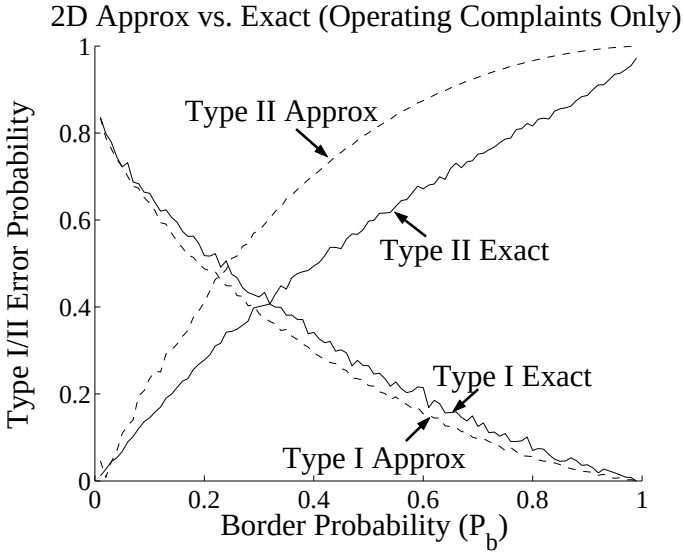

Figure 3.6: Type I/II Errors: $P_i^{approx}, \hat{P}_i^{exact}$ vs. P_{b_i}

In Fig. 3.6, 360 sample points were used to compute the Monte Carlo integrations resulting in the approximate Type I/II Errors, P_i^{approx}, for all border probabilities, P_{b_i}. The grid for P_{b_i} has a resolution of 0.01, and ranges from 0.01 to 0.99. $\hat{P}_i^{exact}$ represents the relative frequency of false alarms and missed detections, for a $100,000$ second simulation run for each grid point, P_{b_i}.

Fig. 3.7 illustrates the 1-norm difference between the exact and approximate Type I (on top) and Type II (on bottom) errors as a function of the number of sub-intervals. Note that as the number of sub-intervals increases from the single interval 2-dimensional approximation to the maximum number possible, the difference between the exact and the approximate methods dies away, and then comes back up slightly. Also superimposed on this graph, as a fixed dotted line, are the multi-dimensional approximation results, for comparison. They are fixed because the result is independent of the number of sub-intervals. One may think of the multi-dimensional result as being most closely related to the result for the maximum number of sub-intervals, because they share the same dimensionality. Results are shown when using both 360 sample points on the left and 3600 sample points on the right, to compare accuracy of the Monte Carlo integration result. Error bars are also shown for all approximations, $\pm$ a single standard error. The points that are circled correspond to the 1-norm Type I/II approximation errors derived from the 2-dimensional alarm system approximation shown as a function of border probabilities in Fig. 3.6.

$$\begin{aligned} \hat{\tau} &= \|\hat{P}_{fa}^{approx} - \hat{P}_{fa}^{exact}\|_1 \triangleq \sum_{i}^{N_{P_b}} |\hat{P}_{fa_i}^{approx} - \hat{P}_{fa_i}^{exact}| \qquad (3.110) \\ SE(\hat{\tau}) &= \sqrt{\sum_{i=1}^{N_{P_b}} \frac{1}{n_i^{approx}} \hat{P}_{fa_i}^{approx}(1 - \hat{P}_{fa_i}^{approx}) + \frac{1}{n_i^{exact}} \hat{P}_{fa_i}^{exact}(1 - \hat{P}_{fa_i}^{exact})} \end{aligned}$$

Eqn. 3.110 illustrates the 1-norm difference and standard error equations, for Type I false alarms. $\hat{P}_{fa_i}$ is the relative frequency and approximate probability of false alarms based upon a large $N = 100,000$-sample simulation. $N_{P_b} = 99$ (the number of grid points that $P_b \in [0, 1]$ spans). n_i represents the number of alarms that sound in the N-sample simulation. To compute the standard error, it is assumed for convenience that alarms are independent, and can be modelled as Bernoulli

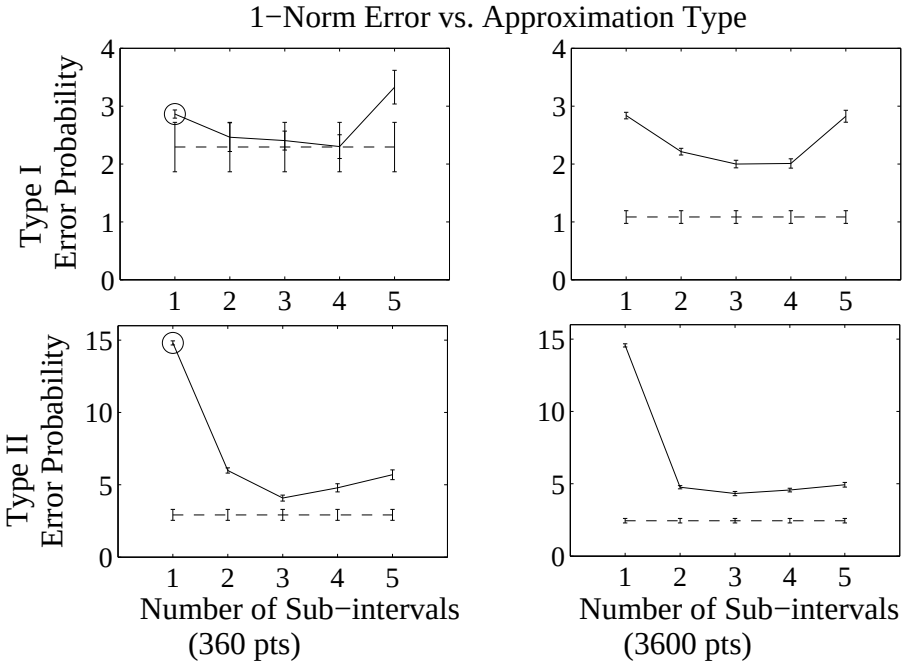

Figure 3.7: Type I/II Approximation Error, $\|\hat{P}^{approx} - \hat{P}^{exact}\|_1$ vs. Number of Sub-intervals

random variables. It is possible to compute all approximations to Type I/II probabilities using the numerical integration method described previously. As such, no direct simulation is required, but in the process of computing the probabilities numerically, Monte Carlo simulation is required. Therefore, the $\hat{P}_{fa}^{approx}$ portion of the standard error is not based on relative frequencies. It is computed numerically, so it should be written as P_{fa}^{approx}. This comes from residual Monte Carlo simulation error, and is added to the standard error equation shown above in an appropriate manner.

3.4.2 Example 2: Svensson Model, Both Arrival and Operating Complaints

For this example, our problem is to design an optimal alarm system that predicts at least one arrival and operating complaint, also defined previously ($1 - P(x_{k+d} < L, x_{k+d+1} < L, \ldots, x_{k+d+m} < L | x_0, \ldots, x_k) \geq P_b$). Recall that we use the same $m = 5$ step ahead prediction

window, and $d = 2$ is the number of steps prior to the start of the beginning of the day. An example of exact vs. approximate Type I/II errors for the 2-dimensional alarm system approximation describing this example as a function of border probabilities is shown in Fig. 3.8.

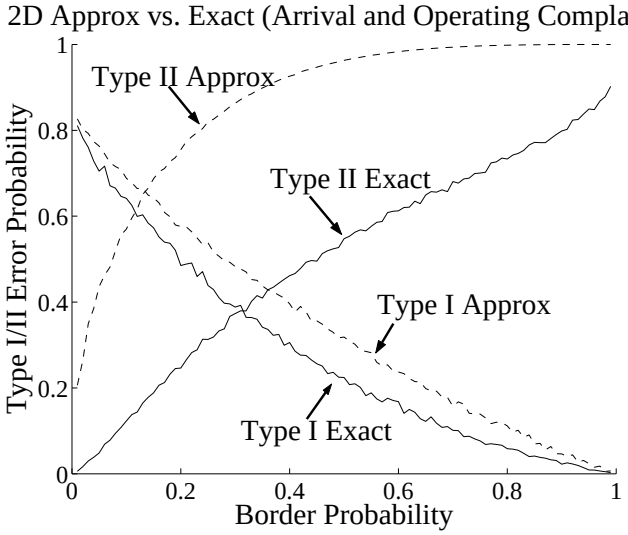

Figure 3.8: Type I/II Errors (Upcrossings): $P_i^{approx}, \hat{P}_i^{exact}$ vs. P_{b_i}

Again, in Fig. 3.8, 360 sample points were used to compute the Monte Carlo integrations resulting in the approximate Type I/II Errors. A $100,000$ second simulation run for the "exact" probability computations was used again as well. Fig. 3.9 shows the 1-norm approximation error as a function of the number of sub-intervals, along with the multi-dimensional approximation results, for comparison. Both Type I/II approximation errors are shown. Results are shown when using 360, 3600, 36000, and 360000 sample points to compare accuracy of the Monte Carlo integration result.

The multi-dimensional approximation method results are shown as fixed dotted lines, where the Type I results are always below the Type II results. Again, the points that are circled correspond to the 1-norm Type I/II approximation errors derived from the 2-dimensional alarm system approximation shown as a function of border probabilities in Fig. 3.8.

The sub-interval approximation results shown have also been computed by making the approximation of each sub-interval with *upcrossings*, as originally defined. However, due to the

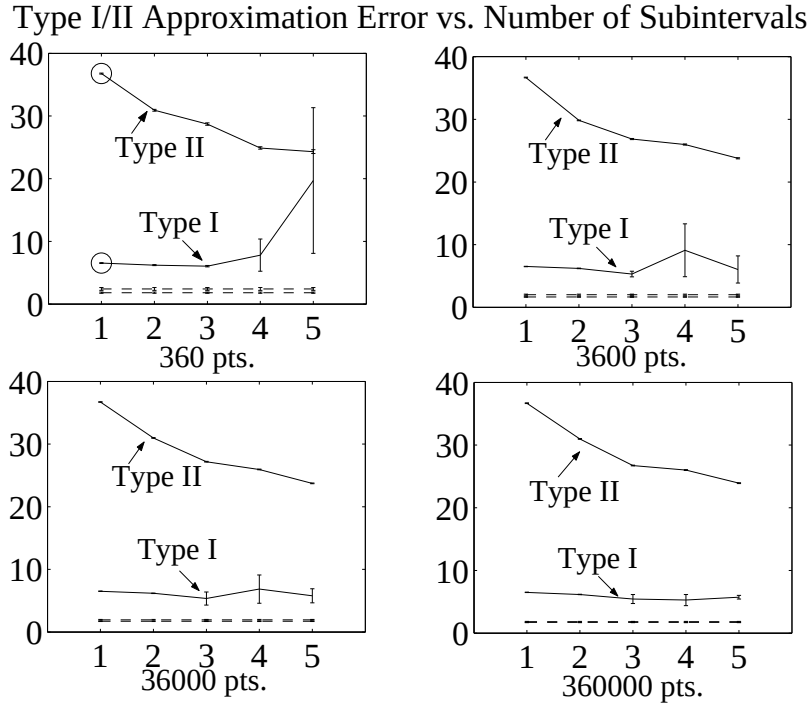

Figure 3.9: Type I/II Approximation Error (Upcrossings), $\|\hat{P}^{interv} - \hat{P}^{exact}\|_1$ vs. Number of Subintervals

nature of the hypothesis being tested, this need not be the sole method of approximation for subintervals. In fact, intuitively arrival complaints can be described more accurately with downcrossings than with upcrossings. Recall that arrival complaints are defined as "exceedances," and late arrivals are allowed. Because exceedances are defined in single time slice, we don't know about the arrival complaint until the exceedance terminates in some subsequent time slice. Therefore, the arrival complaint can best be determined with knowledge of a downcrossing.

Fig. 3.10 illustrates an example of exact vs. approximate Type I/II errors for the 2-dimensional alarm system approximation describing the "downcrossing" version of this case as a function of border probabilities. The same simulation parameters were used as in the "upcrossing" case.

Fig. 3.11 illustrates the revised "downcrossing" sub-interval approximation results. These figures show the Type I/II 1-norm approximation error as a function of the number of sub-intervals,

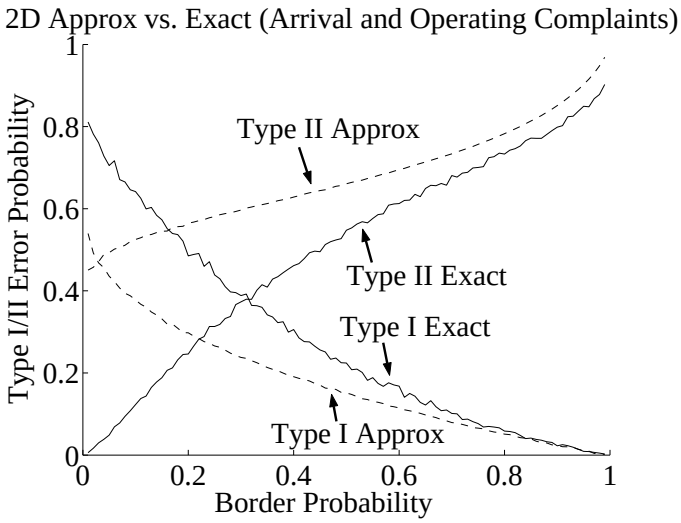

Figure 3.10: Type I/II Errors (Downcrossings): $\hat{P}_i^{approx}, \hat{P}_i^{exact}$ vs. P_{b_i}

along with the multi-dimensional approximation results, for comparison. Here results are shown when using both 360 and 3600 sample points to compare accuracy of the Monte Carlo integration result. Circles correspond to the Type I/II 1-norm approximation error derived from Fig. 3.11.

3.4.3 Discussion of Results for Examples 1 and 2

Before proceeding further, it is necessary to credit work performed by Svensson, [69], p. 94, for his foresight in realizing the need for sub-interval approximations when dealing with a level-crossing in a time interval with non-adjacent endpoints. However, his null hypothesis only deals with a single upcrossing in the entire time interval, as opposed to "at least one," as we do here. Therefore, even though the approximations that are being used in both cases are identical, the fundamental questions being asked, or null hypotheses, are quite different. Consequently, the resulting alarm system characteristics and will be different as well. We'll discuss and compare the merits of the multiple sub-interval approximation method and the multi-dimensional approximation, which is tailored for the case when we have multiple level crossings over a time interval.

The detection probabilities, or Type I/II error probabilities will be biased when using a

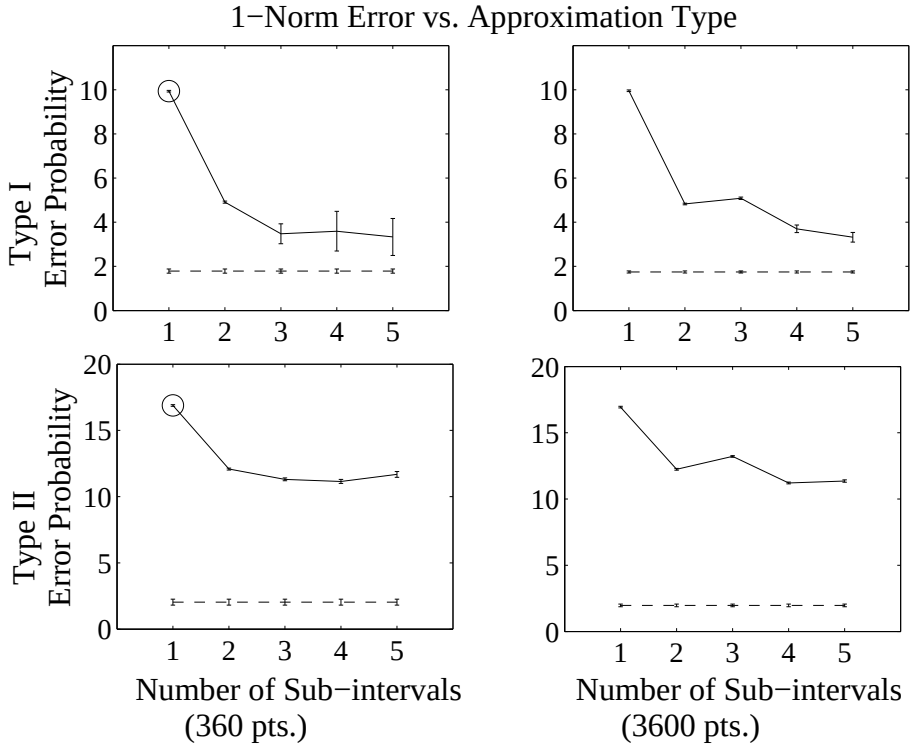

Figure 3.11: Type I/II Approximation Error (Downcrossings), $\|\hat{P}^{interv} - \hat{P}^{exact}\|_1$ vs. Number of Sub-intervals

sub-interval approximation to Svensson's hypothesis. When using the single interval 2-dimensional approximation, the detection probability will be higher than for the exact event. Using multiple sub-interval approximations will result in the detection probability being lower than for the exact event. Different biases will apply for our hypothesis. Recall from Fig. 3.7, that when the number of sub-intervals increases, the error between the exact and the approximate methods dies away, and then comes back up slightly. This phenomenon is related to the competing effects of improvement in approximation with increased number of subintervals, and introduction of errors from two sources.

The first source of error is due to the compounding "rectangularization" approximations that are made with each new 2-dimensional sub-interval. The second source of error is due to the approximation made for the entire exact alarm region, by using the union of disjoint sub-intervals. The latter is the primary source of error, and has to do with the fact that the well-defined region of

integration, $\Omega_{A_{approx}}$, is a heuristic approximation of $\Omega_{A_{exact}}$.

Fundamentally, these two alarm regions are not the same. Recall that for sub-interval approximations, $A_{approx} = \bigcup_{i=1}^{N_s} A_i$, which is simply an approximation of A_{exact}. There is no guarantee that this approximation will improve with the number of subintervals. In fact, intuitively we know that sub-interval approximations capture only the number of "upcrossings" in all sub-intervals. The exact alarm condition will capture more complicated cases which don't restrict upcrossings to occur in specific sub-intervals. Recall the definition of the exact alarm condition for operating complaints only:

$$
\begin{aligned}
C_{exact} &= \{\mathbb{X} \in \Omega_{C_{exact}} \subset \mathbb{R}^{m+1}\} \\
&= \neg\{x_k < L, x_{k+1} < L, \ldots, x_j < L, \ldots, x_{k+m} < L\} \setminus \{x_k > L\} \\
&= \overbrace{\neg}^{\text{not}} \underbrace{\{x_k < L, x_{k+1} < L, \ldots, x_j < L, \ldots, x_{k+m} < L\}}_{\text{no complaints}} \overbrace{\setminus}^{\text{take away}} \underbrace{\{x_k > L\}}_{\text{arrival complaints}}
\end{aligned}
$$

Here we are saying that operating complaints can be defined with any sequence of process values above or below the critical threshold, as long as **all** of the values don't lie above the threshold, and the first process value is not above the threshold (to prevent us from counting arrival complaints). An example of where the sub-interval approximation fails is when upcrossings are checked for each pair of adjacent sub-intervals. As we increase the number of sub-intervals, this restriction becomes tighter and tighter until we reach the maximum number of sub-intervals, and each pair of adjacent sub-intervals is identically each pair of adjacent time slices. In this case (as in all cases), the process value is restricted to be above the critical threshold at the last time slice. However, we know in the case of the exact alarm condition, there is no such restriction at the last time slice. Furthermore, geometrically speaking, we know that the space carved out in $m+1$-dimensional space by $\Omega_{A_{approx}}$ will never be the same as the space $\Omega_{A_{exact}}$.

As we've seen in Fig. 3.7, the multi-dimensional approximation method (dotted line) has

the lowest error of all methods, regardless of the number of subintervals, or type of error probability. In fact, as we increase the number of sampling points, the error naturally decreases, but even further below the sub-interval method. It is obvious that using the multi-dimensional method provides for better accuracy than the 2-dimensional or the multiple sub-interval approximation methods. This is further evidenced by Example 2, in which we compare the multi-dimensional methods to two different kinds of sub-interval methods: both upcrossings and downcrossings in Figs. 3.9 and 3.11.

Notice also that for each diagram, the error bars naturally decrease in size with the number of sampling points chosen for the numerical integration. This is most evident in Fig. 3.9. Here, there is much difficulty in reducing the error in the numerical integration when the number of sub-intervals is large for the Type I error probability. However, with a sufficient number of points used in the numerical integration routine, they can be reduced. Part of the reason for this anomaly may be due to the fact that using upcrossings to approximate the exact alarm condition for both arrival and operating complaints is probably not as accurate as using downcrossings. This is evident in Fig. 3.11. Notice that the magnitude of the error and the error bars for downcrossings is much less than for upcrossings, particularly for Type II errors. However, the main conclusion to gather from these results is the fact that the multi-dimensional method is always better with respect to accuracy of the approximation.

From an engineering and alarm design standpoint, we now need to consider the trade off between accuracy and computational attractiveness. From the standpoint of the multi-dimensional approximation method, it has the same computational burden as using the sub-interval method when $N_s = m$. Otherwise if $N_s < m$, the sub-interval method has a more desirable computational load. Recall that N_s is the number of subintervals, and m is the number of steps in the prediction window. Therefore, N_s can be considered an alarm design parameter. We are required to use m terms when implementing the multi-dimensional approximation method. However, the accuracy is better.

Reduced design time can be achieved by using the sub-interval method for the number of

sub-intervals that yields the lowest error, trading off accuracy. The judgement of which method to use is left to the discretion of the designer. When using 360 points in the numerical integration scheme for operating complaints only, and for the Type I error probability, using 4 sub-intervals is probably as good a choice or perhaps better than using the multi-dimensional method. Here the accuracy is about the same, and the design time will be a little less due to the computation burden of having 2^4 terms to evaluate in lieu of 2^5.

Recall that one of the issues mentioned before was trying to make the multi-dimensional method more efficient computationally. An alternative to using the inlcusion/exclusion rule, which has $\mathcal{O}(2^m)$ terms, was to compute its complement instead, i.e. $P(\bigcup_i A_i) = 1 - P(\bigcap_i A_i')$. This method seems like it should turn out to be more useful here than for sub-intervals. Using it for sub-intervals requires $\mathcal{O}(2^{2^{N_s}})$ terms, in lieu of the $\mathcal{O}(2^{N_s})$ terms required by computing $P(\bigcup_i A_i)$ directly. For the multi-dimensional approximation, $\mathcal{O}(m)$ terms are required, in lieu of the $\mathcal{O}(2^m)$ terms needed by using the inclusion/exclusion rule. Therefore, using this method for the multi-dimensional approximation will considerably reduce the number of terms to compute.

However, there are still instances in which using the inclusion/exclusion rule for the multi-dimensional method provides similar accuracy for a smaller number of Monte Carlo integration sample points. In other words, even though the inclusion/exclusion rule requires $\mathcal{O}(2^m)$ terms, the numerical integration error for all of the terms will be much smaller than the resulting error when using the same number of Monte Carlo integration sample points for the alternate method. Even though the computation requires an order of magnitude fewer terms, computing the probability with the same accuracy takes just as long. The number of sample points required to achieve the same accuracy is much greater, hence the computational burden is the same or worse. This intuitive reason for this is that the complement of the events in question are located in a section of the probability space that is harder to sample by using Monte Carlo integration than the original events. Therefore, although computing the probabilities requires fewer terms, it takes just as long to attain the same level of accuracy.

A picture of the multi-dimensional approximation results vs. the sub-interval and exact methods is shown in Fig. 3.12. The multi-dimensional approximation and the sub-interval methods have used the same number of numerical integration sample points (3600) for this figure as well.

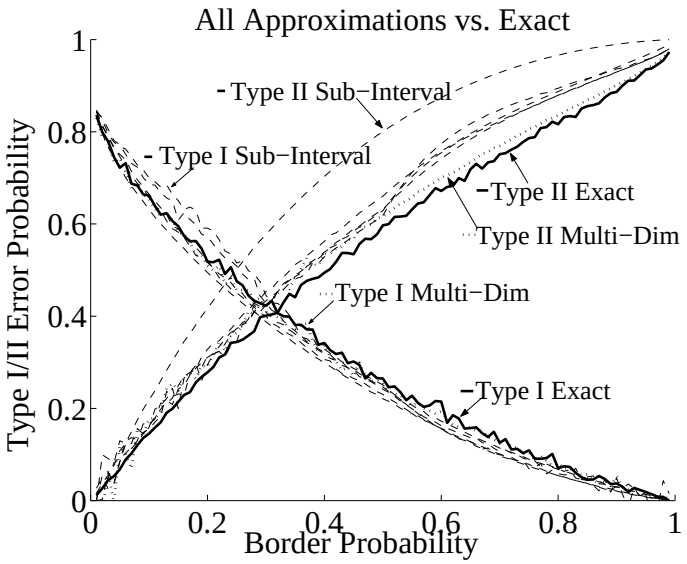

Figure 3.12: Type I/II Error Comparisons, Operating Complaints Only

Fig. 3.12 is the case for Example 1, operating complaints only. Notice that there is trend as the number of sub-intervals (dashed lines) is increased (up to a point), and how closely the multi-dimensional method (dotted lines) track the exact results (thick solid lines), which is more evident for Type II errors. There is an apparent convergence to the exact results, which is also more evident for Type II errors.

As a prelude to presentation of the next 2 examples, we may conclude that any approximation with desirable accuracy/computational tradeoff characteristics may be chosen as the basis for designing optimal alarm systems. This is preferable to performing a comparison between the exact method and all of the different approximations, as shown in Figs. 3.7, 3.9, and 3.11. We've seen that the multi-dimensional approximation is the best in terms of accuracy, even if it does amount to a modest amount of computing time. We will develop an engineering judgement as to the best

estimate of the number of subintervals needed for the lowest Type I/II error probability residuals, if we decide against using the multi-dimensional approximation. Our purpose is for design, so additional accuracy is at the expense of the designers' time. We are not using these results for real-time control where a digital signal needs to be computed and delivered within certain prespecified timing specifications.

Furthermore, running the exact method for comparison in the next 2 examples would require an extreme amount of computational effort due to the use of the "counting" method, where the statistics of Type I/II error probabilities are obtained by using ergodicity (See DeMaré [15] for further details on this). This is especially true due to the fact that very few events will be generated purely from simulation due to the nature of the spectral content of the processes in question, and their relative means. The simulation would have to run an extremely long time in order to generate results for the exact method that could feasibly be compared with the approximate methods.

3.5 Results for Examples 3 and 4

3.5.1 Example 3: Federspiel Model, Operating Complaints Only

As shown in Fig. 3.13, results for the multi-dimensional method (the dotted lines) can be compared to the sub-interval method (the solid lines). For this example, we use the complaint model and relevant parameters, as specified for operating complaints only. Recall that we use a 12-step ahead prediction window, or $m = 12$. This corresponds to a 4-hour period in which operating complaints can occur either during the morning or afternoon period of the day. In Fig. 3.13, convergence with sub-intervals (solid) to the multi-dimensional method (dotted) is apparent, so therefore we can almost guess as to where the exact Type I/II error probabilities as a function of P_{b_i} might lie.

Notice that we only show the graph's border probability out to 0.65 because of a numerical issue beyond that point. This is due to a very low alarm probability beyond that point (it is close

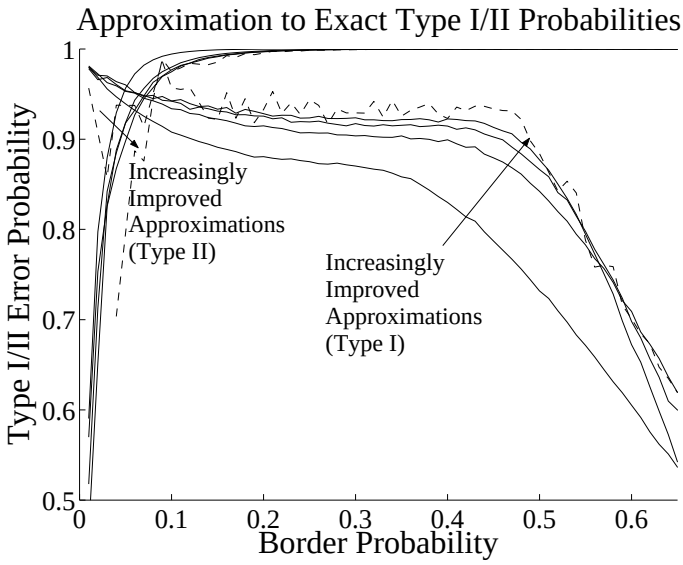

Figure 3.13: Type I/II Errors for Operating Complaints Only

to zero). The probability of false alarms is scaled by this quantity, hence resulting in numerical problems. All approximations shown have used 3600 numerical integration sample points, and only the first 4 sub-intervals are shown in solid lines.

3.5.2 Example 4: Federspiel Model, Both Arrival and Operating Complaints

For the last example, we use the complaint model and relevant parameters, as specified for both arrival and operating complaints. Recall that we use the same $m = 12$ step ahead prediction window as in Example 3, and $d = 3$ as the number of steps prior to the start of the day. In this 1-hour period prior to the beginning of the day, we want to predict both arrival and operating complaints in the ensuing 4-hour morning period. As shown in Figs. 3.14 and 3.15, results for the multi-dimensional method (dotted lines) can be compared to the sub-interval method (solid lines) when using upcrossings or downcrossings (Figs. 3.14 and 3.15, respectively).

As in Example 3, Fig. 3.14 illustrates convergence of the sub-interval method (solid) to the

multi-dimensional method (dotted). Again, we can "guess" as to where the exact Type I/II error probabilities as a function of P_{b_i} might lie. For Fig. 3.15, there is barely a perceptible difference among the different sub-interval results and multi-dimensional result shown, barring low values of P_{b_i}.

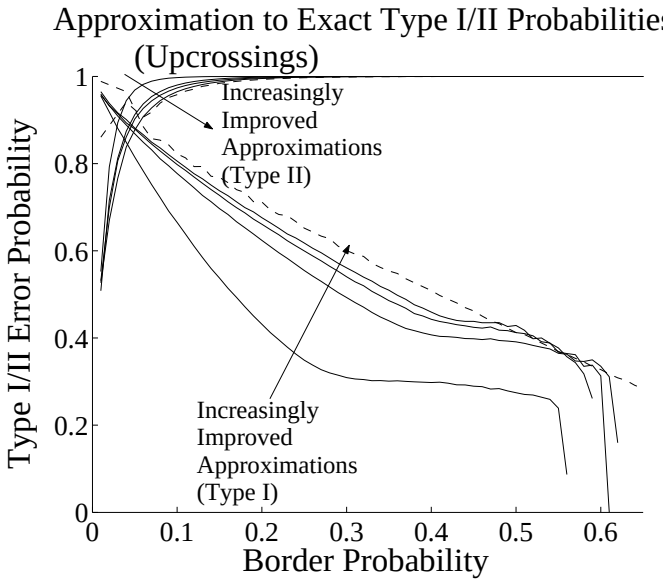

Figure 3.14: Type I/II Errors for Both Arrival & Operating Complaints

All approximations shown have used 3600 numerical integration sample points. For Fig. 3.14 (upcrossings), the first 4 sub-intervals are shown. The first 3 sub-intervals are shown in Fig. 3.15 (downcrossings).

3.5.3 Discussion of Results for Examples 3 and 4

For examples 3 and 4, as the number of sub-intervals increases, the number of terms that need to be computed increases exponentially as $\mathcal{O}(2^{N_s})$. Therefore the error involved with the numerical integration for each of this terms scales respectively. As a result, the noise will increase with the number of sub-intervals. Similarly, the noise will also be even greater for the multi-dimensional approximation method, which requires $\mathcal{O}(2^m)$ terms ($m = 12$). This is mainly evident

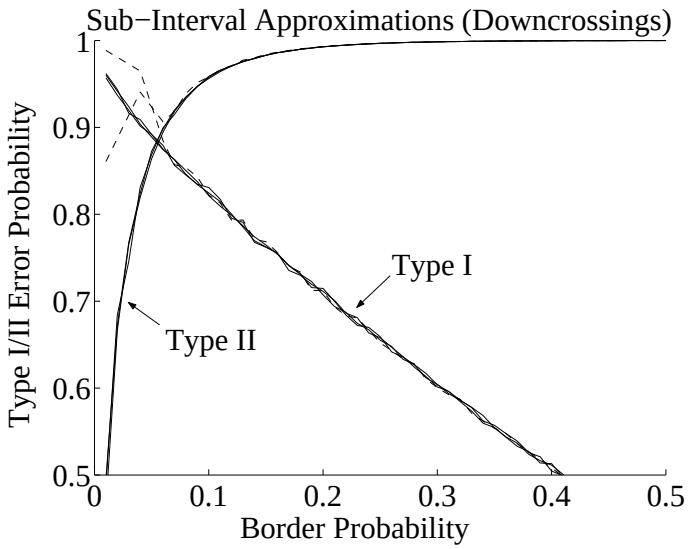

Figure 3.15: Type I/II Errors for Both Arrival & Operating Complaints

in Fig. 3.13. Figs. 3.14 and 3.15 demonstrate that the noise due to the multi-dimensional method (dotted lines) is most prevalent for low values of P_{b_i}, otherwise there is very little difference between this and the sub-interval results (solid lines).

Recall the number of numerical integration sample points used for the multi-dimensional method (3600). This was quite a large number, and took significant design time. Improved accuracy and reduced noise can be achieved by using an even larger number of sample points, at the expense of an increased computational burden. The same logic holds for increasing the number of sub-intervals, however, potentially only up to a point. Recall the results and discussion from Examples 1 and 2 that provided the rationale for the behavior of the error when the number of sub-intervals increases. In those examples, the error between the exact and the approximate methods dies away, and then comes back up slightly. However, as seen in Examples 3 and 4, only a few sub-interval results are required for convergence to the exact results shown in Figs. 3.13, 3.14, and 3.15. Therefore, the same behavior of the error between the exact and the approximate methods dying away, and then coming back up slightly is not exhibited with the given limited range of sub-interval results.

We can conclude that for Examples 3 and 4, using the sub-interval method is probably a better approximation to use both in terms of computational burden and accuracy. When using just a few sub-intervals ($N_s = 3$ or 4), we get reasonably convergent behavior, with a modest number of terms to compute and less accrued integration error $\mathcal{O}(2^{N_s})$. However, when using the same number of integration sampling points for the multi-dimensional approximation, the computational burden is high, $\mathcal{O}(2^m)$, where $m = 12$. Therefore, not only will this method result in a greater computational burden, but it will also accrue more error due to the number of terms, as manifested in the noise evident in Fig. 3.13 and Figs. 3.14, 3.15 for low values of P_{b_i}.

Furthermore, it seems that in the case of Example 4 specifically, when modelling both arrival and operating complaints with downcrossings as opposed to upcrossings, the overall approximation is much better (Fig. 3.15). It is most evident in how close the sub-interval method comes to the multi-dimensional method in this Figure as opposed to Fig. 3.14. More sub-intervals might be required to achieve an improved approximation to the multi-dimensional result in Fig. 3.14, which we know will be closest to the exact result. In fact, we can see from Fig. 3.15, that it is very likely that applying sub-interval method for the fewest number of sub-intervals is sufficient for design, and in this particular case, provides better accuracy as well as computational efficiency than the multi-dimensional method.

Therefore, Fig. 3.15 provides us with further evidence to our intuition that it is more appropriate to model the case of both arrival and operating complaints with downcrossings than with upcrossings, when using the sub-interval method. There is another important observation to note in the use of upcrossings vs. downcrossings, independent of the hypothesis being tested (operating vs. both arrival and operating complaints). In Figs. 3.13, and 3.14, both hypotheses being tested use upcrossings to approximate the alarm region for differing null hypotheses. In these two figures, we see a distinct similarity in qualitative appearance. Namely, there is a change in slope at a critical border probability point whereas there is no such point for the downcrossing model in Type I errors . This critical point may in fact be related to the optimal selection criteria for a new

alarm design metric to be introduced in a future chapter.

3.6 Conclusion

Considering the results from all of the 4 examples presented, a general conclusion can be made concerning the use of the approximations detailed in this chapter. When the null hypothesis characterizes multiple events over a large prediction window, we are probably better off using the sub-interval approximation, and a small number of sub-intervals. However, for cases in which the prediction window is more modest, the multi-dimensional approximation will not only provide better accuracy, but the computational burden will tend to be on par with the sub-interval method.

Nothing yet has been mentioned about the actual alarm design procedure or results from its implementation, which can be performed by selecting the optimal P_b value. All of the theory presented in this chapter hinges on a fundamental concept from decision theory. The Neyman-Pearson Lemma provides us with this decision framework. This lemma allows us to design an optimal alarm system that will elicit the fewest possible false alarms for a fixed detection probability. The fixed detection probability is of course parameterized by the border probability, P_b. This border probability therefore serves as a free parameter, and hence as the primary design metric.

The steps required for design of an alarm system have been covered in previous work [69, 70]. Suggestions for choosing an optimal P_b value in the design of the alarm system were provided here as well. However, they tend to be based either on purely heuristic trial-and-error approaches, or cost functions. For cost functions, sometimes it is easy to assign particular costs to events that penalize the probability of alarms, false alarms, and missed detections, etc. Assigning these costs requires heuristic knowledge of the risk-reward tradeoff in terms of Type I/II errors. Since the events that typically need to be predicted for simple non-optimal alarm systems are extremely critical, Type I errors (probability of false alarms) are obviously much less costly than Type II errors (probability of missed detections). Therefore, because the costs are well known, simple alarm systems of the type

described in the introduction are designed for Type II error avoidance.

However, an optimal alarm system can also be designed for very critical events in lieu of a simple alarm system. In fact, the entire optimal alarm literature focuses on events of a critical or catastrophic nature. In this case, we would still have a very good idea of the costs of Type I/II errors. As such, these costs could easily be used to act as weights in the cost function described above in order to choose an optimal P_b value in the design of an optimal alarm system.

There is an entire field, and a ubiquitous literature dedicated to the study of issues concerning tradeoffs as discussed above known as decision theory and hypothesis testing (See VanTrees [73] as an example). The purpose of this chapter was not to provide any new contributions to this field, but rather to use existing theory to our advantage. Recall from our introduction that the main incentive for considering optimal alarm system design was to address the needs of the engineering system we've described as Examples 3 and 4 for multiple complaint prediction, involving events of a less severe nature. Here, we require a more subtle balance between Type I and Type II errors.

In this case, a more sophisticated optimal alarm system is of interest, but as we've seen, the results tend to show very high Type II errors over a large range of border probabilities. Because of this, no matter what value of P_b is selected, there will always be a high probability of a "missed detection." Looking at figures in the previous results section, we see that even when the value of P_b is low, resulting in a smaller probability of missed detection, the tradeoff is that the false alarm probability is quite high. It seems as if we can never avoid almost certain missed detection in order to achieve a reasonable false alarm probability.

Furthermore, little is known about the costs of Type I and Type II errors for our particular application. However, as a general rule of thumb in building operations we are probably much better off missing a possible customer complaint and saving the money in labor hours and dollars charged to respond, than anticipating and responding to a complaint that will never happen. In this case a facility manager might actually perform some sort of preemptive action in response to a false alarm, causing even more actual complaints. So in a sense, false alarms have a higher cost in our particular

application, unlike most applications where critical or catastrophic events might occur, and where missed detections are much more costly. Therefore, we may benefit from a more creative selection of an optimal P_b value, focused on optimizing false alarms.

There are several other decision rules that can be used from hypothesis testing/decision theory, in lieu of the Neyman-Pearson decision rule. Among these are Bayesian, MAP (maximum a posteriori), maximum likelihood, and the minimax criteria. The latter is derived from the Bayesian criterion, and seeks to minimize the maximum risk. Some recent interesting developments have even described adaptive on-line techniques using the Bayesian formulation [2]. However, there are still considerable computational issues, and a well-defined cost function is still required, even when the posterior probability is adaptively updated.

We have parameterized the Type I/II error probabilities based upon the Neyman-Pearson decision rule via the border probability design metric P_b. If we choose to create a cost function that weights the Type I/II errors equally, then we have suitable candidate for the minimax choice as described above. In fact, we can read the optimal P_b value right off of the figures in the previous section by applying this minimax decision rule. However, if we use this rule for Examples 3 and 4, the results still tell us that both the probability of false alarms and missed detections will still be quite high (close to .9 or in some cases higher).

For the application presented in Examples 3 and 4, we don't seem to have much control over the probability of missed detection without severely compromising the probability of false alarms. As such, we may choose a cost function that has its roots in the accumulation of multiple complaints. In other words, since lowering the probability of missed detections is difficult, we can focus on minimizing false alarms with judicious selection of the cost function, since that seems to have a higher priority for our application. Minimizing false alarms achieves the same objective as maximizing correct alarms, and since we're dealing with **multiple** complaints, the objective can be reformulated even further to maximizing the expected number of complaints conditioned on an alarm, $E[C|A]$. Unlike the probability of false alarm, this can be expressed as a convex, rather

than a monotonic function of the border probability, and a unique maximum can be found. It is theoretically an excellent alternative to using a cost function in finding the optimal value of P_b, when applying the Neyman-Pearson lemma. The theory and practical merits of this new metric will be covered in detail in Chaps. 5 and 7. We will find that using the $E[C|A]$ design method has favorable computational properties, because finding the metric as a function of P_b has a computational burden that scales linearly in the number of time steps under consideration. Furthermore, in Sec. 7.2, we will revisit the practical examples introduced in this chapter, to establish the practical usefulness of using a particular design criteria for optimal alarm design.

Chapter 4

Discretization of Level-Crossing Formulae

4.1 Introduction

This short chapter can be viewed as an independent contribution to applied mathematical statistics, based upon results from level-crossing theory. Although at first read the chapter may seem unrelated to the remainder of the thesis, the results will be used in subsequent chapters. The most important concept in this chapter is the introduction of approximations of stationary and non-stationary Gaussian random process level-crossing formulae. These equations come from decades-old theory, derived mainly from the works of Rice [61], Cramér and Leadbetter [14].

The central idea is to use discrete approximations in lieu of their counterpart continuous formulae. We will need to extend the discrete-time theory presented thus far, and formulate additional discrete-time theory, both which heavily rely on level-crossing formulae. We must make sure that all level-crossing equations for Gaussian random sequences to be used are either in discrete form, or are discretized versions of continuous equations. This is necessary in order for the application of

theory to fit the discrete-time paradigm it is presented in. The equations of interest may be useful for both optimal alarm design and optimal control in future chapters. The level-crossing formulae of interest quantify the expected number of upcrossings over a specified time interval. These formulae have representations in both continuous and discrete time. Several discrete time formulae have been introduced in [7, 38] by Kedem and Barnett. Continuous formulae have mainly been seen in Rice [61], Cramér and Leadbetter [14].

We may easily choose to use the discrete-time formula paradigm of Kedem and Barnett's work, [7, 38] however, it can be shown for specific cases that the resulting formulae are not in closed form, even though they are exact. We prefer to have closed-form solutions when dealing with real-time control for a number of reasons. In real-time control systems, the digital control signal needs to be computed and delivered within tight timing specifications. Numerical solutions are required for the discrete-time level-crossing formulae, and therefore incurs a greater computational burden. As a result, these discrete-time formulae are not conducive to delivering a computation at requested time for digital control purposes. For optimal alarm design, using closed form solutions of level-crossing formulae are still preferable, due to the decreased computational burden and design time. To summarize, instead of using generalizations of the discrete-time level-crossing formulae, we perform very straightforward discretizations of the continuous-time level crossing formulae. A flowchart illustrating the basic problem is shown in Fig. 4.1.

4.2 Theory

4.2.1 Zero-Level Upcrossings of a Zero-Mean Stationary Gaussian Random Sequence

The expected number of zero-level upcrossings of a zero-mean stationary Gaussian random sequence in a *single* time interval is provided in Eqn. 4.1, in discrete time [38].

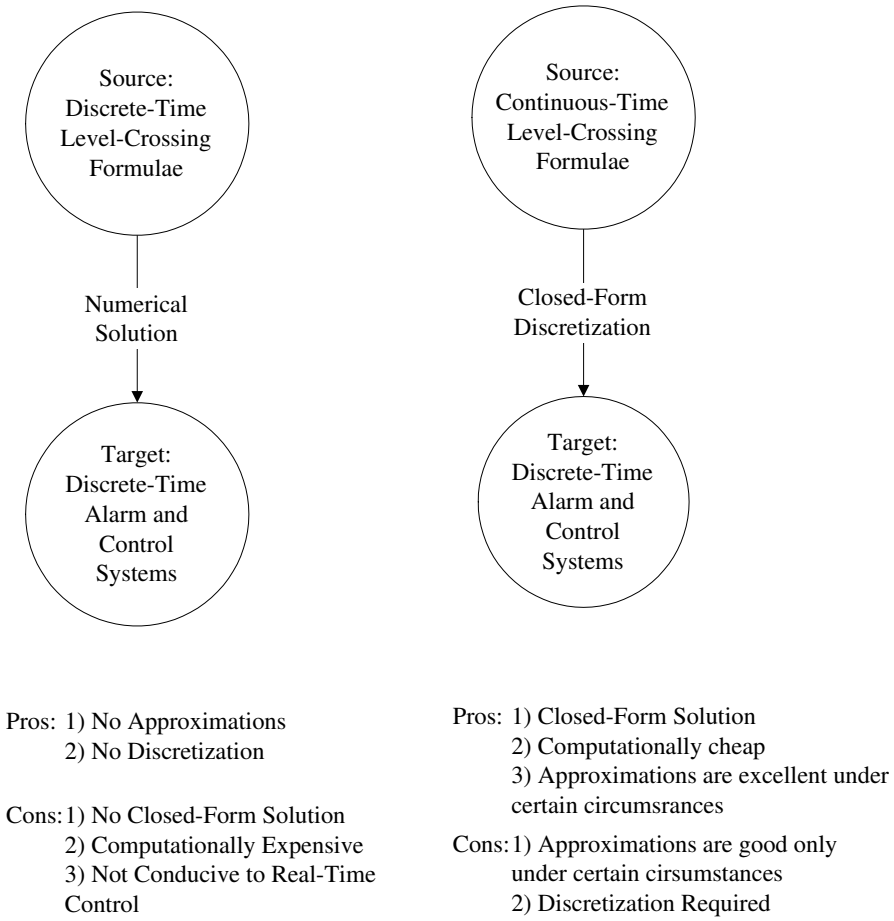

Figure 4.1: Comparative Flowchart

$$
\begin{aligned}
E[C] &= P(x_k < L, x_{k+1} > L) \qquad (4.1)\\
&= \frac{1}{4} - \frac{1}{2\pi}\arcsin(\rho_1)\\
\text{when L} &= 0\\
\text{and } \mu_x &= 0\\
\text{where } \rho_1 &\triangleq \frac{cov(x_{k+1}, x_k)}{\sqrt{Var(x_{k+1})Var(x_k)}}
\end{aligned}
$$

For multiple time intervals, Eqn. 4.1 can simply be multiplied by the number of time

intervals in question, m. Therefore, we get

$$E[C] = mP(x_k < 0, x_{k+1} > 0) = \frac{m}{4} - \frac{m}{2\pi}\arcsin(\rho_1) \tag{4.2}$$

Notice that this formula has a very nice closed form. It can easily be parameterized by the one-step autocorrelation, ρ_1. Therefore, it is not necessary to discretize to obtain a closed-form. However, for clarity of presentation, we'll use this simple case to illustrate how the expected number of upcrossings quantified by Rice's continuous time formula [61] can be discretized. We know that the standard level-crossing problem yields the mean upcrossing frequency, ν_x, given by Rice's famous formula [14, 19, 61], Eqn. 4.3.

$$\nu_x = \frac{\sigma_{\dot{x}}}{2\pi\sigma_x} e^{-\frac{1}{2}\frac{(L-\mu_x)^2}{\sigma_x^2}} \tag{4.3}$$

We can also compute the expected number of upcrossings, by taking the product of this mean upcrossing frequency and the time period in question, t. Therefore, this gives us an expression for $E[C]$, shown in Eqn. 4.5.

$$E[C] = \nu_x t \tag{4.4}$$

$$= \frac{t}{2\pi}\frac{\sigma_{\dot{x}}}{\sigma_x} e^{-\frac{1}{2}\frac{(L-\mu_x)^2}{\sigma_x^2}} \tag{4.5}$$

However, notice that this equation is a continuous time representation of the process, $x(t)$, rather than its discrete-time equivalent, x_k. The discrete time case for expected number of zero-crossings has seen a comprehensive review, [7, 38], in which an analogous cosine formula is derived easily from Eqn. 4.2. This formula is shown in Eqn. 4.6.

$$\rho_1 = \cos\left(\frac{2\pi E[C]}{m}\right) \tag{4.6}$$

The derivation of Eqn. 4.2 can easily be extended to the case for when we are considering non-zero level upcrossings, but the cosine formula, Eqn. 4.6, will no longer hold. Let's now return

to the discretization of Eqn. 4.5, and assume that the time interval in question is $t = mT_s$, where T_s is the sampling interval for the discretized system.

$$\begin{aligned} E[C] &= \frac{mT_s}{2\pi}\frac{\sigma_{\dot{x}}}{\sigma_x}e^{-\frac{1}{2}\frac{(L-\mu_x)^2}{\sigma_x^2}} \\ &= \frac{mT_s}{2\pi}\frac{\sigma_{\dot{x}}}{\sigma_x} \end{aligned} \tag{4.7}$$

$$\begin{aligned} \text{when L} &= 0 \\ \text{and } \mu_x &= 0 \end{aligned}$$

$$\begin{aligned} \text{Recall } \dot{x}(t) &= \frac{dx}{dt} = \lim_{\Delta t \to 0}\frac{x(t+\Delta t)-x(t)}{\Delta t} \\ &\approx \frac{x(t+\Delta t)-x(t)}{\Delta t} \\ &\approx \frac{x_{k+1}-x_k}{T_s} \\ \Rightarrow \sigma^2_{\dot{x}(t)} = Var(\dot{x}(t)) &\approx Var\left(\frac{x_{k+1}-x_k}{T_s}\right) \end{aligned} \tag{4.8}$$

This is essentially a first-order truncation of a Taylor series approximation to the derivative, or equivalently, a "forward" difference approximation.

$$\begin{aligned} \Rightarrow Var\left(\frac{x_{k+1}-x_k}{T_s}\right) &= \frac{1}{T_s^2}\left[Var(x_{k+1})+Var(x_k)-2\rho_1\sqrt{Var(x_{k+1})Var(x_k)}\right] \\ \text{Recall } Var(x_{k+1}) &= Var(x_k) \text{ for stationary processes} \\ \Rightarrow \frac{\sigma_{\dot{x}(t)}}{\sigma_{x(t)}} &\approx \sqrt{\frac{2(1-\rho_1)}{T_s^2}} \end{aligned}$$

This same result can be found be using a "backward" difference in lieu of a "forward" difference or first-order Taylor series approximation to the first derivative. These type of approximations can only be used in conjunction with the one-step autocorrelation, ρ_1. Although more accurate than these approximations by $\mathcal{O}(T_s^2)$, a "central" difference cannot be used for comparison, because it would require parametrization by the two-step autocorrelation, ρ_2, instead of the one-step autocorrelation, ρ_1.

From Kedem [38], on pg. 129-130, the discrete and continuous time formulas were shown to be related simply by characterizing the discrete time steps via a difference, Δ, and taking the limit

as $\Delta \to 0$. Therefore we can similarly prove sufficiency here such that $\lim_{T_s \to 0} \sqrt{\frac{2(1-\rho_1)}{T_s^2}} = \frac{\sigma_{\dot{x}(t)}}{\sigma_{x(t)}}$. The derivation is provided as follows:

$$\begin{aligned}
\text{Note } \rho_1 &= \rho(T_s) \\
\text{Recall } \lim_{T_s \to 0} x^2(T_s) &= \left[\lim_{T_s \to 0} x(T_s)\right]^2 \\
\Rightarrow \lim_{T_s \to 0} x(T_s) &= \sqrt{\lim_{T_s \to 0} x^2(T_s)} \\
\Rightarrow \lim_{T_s \to 0} \sqrt{\frac{2(1-\rho_1)}{T_s^2}} &= \sqrt{\lim_{T_s \to 0} \frac{2(1-\rho(T_s))}{T_s^2}}
\end{aligned}$$

Now use L'Hôpital's rule twice:

$$\begin{aligned}
\Rightarrow \lim_{T_s \to 0} \sqrt{\frac{2(1-\rho_1)}{T_s^2}} &= \sqrt{\lim_{T_s \to 0} \frac{-\rho'(T_s)}{T_s}} \\
&= \sqrt{\lim_{T_s \to 0} -\rho''(T_s)} \\
&= \sqrt{-\rho''(0)} \\
&= \sqrt{-\frac{R''_{XX}(0)}{R_{XX}(0)}} \\
&= \sqrt{\frac{\sigma^2_{\dot{x}(t)}}{\sigma^2_{x(t)}}} \\
&= \frac{\sigma_{\dot{x}(t)}}{\sigma_{x(t)}}
\end{aligned}$$

where the 2^{nd} to last equality can be shown by using the assumption that the relevant spectral moments exist, as on pg. 14 of Kedem [38]. A summary of the approximation performed by discretizing the continuous formula is as follows:

$$\begin{aligned}
\text{Recall } E[C] &= \frac{mT_s}{2\pi}\frac{\sigma_{\dot{x}}}{\sigma_x} \\
&\approx \frac{mT_s}{2\pi}\sqrt{\frac{2(1-\rho_1)}{T_s^2}} \\
&= \frac{m\sqrt{2(1-\rho_1)}}{2\pi}
\end{aligned}$$

So for the case of a single time interval, where we choose $m = 1$ without loss of generality,

we have the discretization of Rice's continuous formula, $E[C] \approx \frac{\sqrt{2(1-\rho_1)}}{2\pi}$ that approximates the discrete-time formula $E[C] = \frac{1}{4} - \frac{1}{2\pi}\arcsin(\rho_1)$. The approximation is best when ρ_1 is closest to 1, as seen in Fig. 4.2.

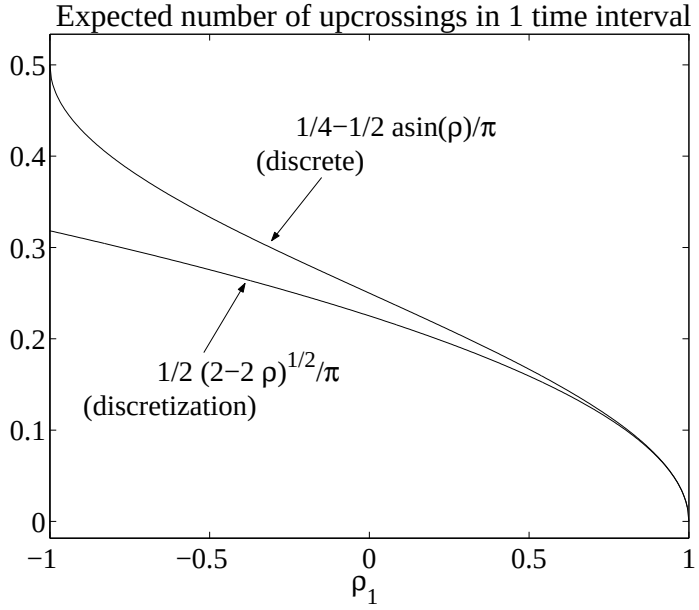

Figure 4.2: Discretization vs. Exact Discrete Level Upcrossings

4.2.2 Level Upcrossings of a Stationary Gaussian Random Sequence

From a more general standpoint, Eqn. 4.1 from the previous section can easily be extended to the case of a non-zero level and non-zero mean , as shown in Eqns. 4.9-4.11. This represents the exact discrete formula for the expected number of upcrossings of a stationary Gaussian random sequence. Here there is no restriction on the level or the mean of the process, which can both be non-zero.

$$
\begin{aligned}
E[C] &= \sum_{i=0}^{m-1} P(x_{k+i} < L, x_{k+i+1} > L) && (4.9)\\
&= \sum_{i=0}^{m-1} \int_{-\infty}^{L} \int_{L}^{\infty} \mathcal{N}(\mathbf{x}_i; \mu_{\mathbf{x}_i}, \Sigma_{\mathbf{x}_i}) d\mathbf{x}_i \\
\text{where } \mathbf{x}_i &= \begin{bmatrix} x_{k+i} \\ x_{k+i+1} \end{bmatrix} \in \mathbb{R}^2 \\
\mu_{\mathbf{x}_i} &= \begin{bmatrix} \mu_{x_k} \\ \mu_{x_k} \end{bmatrix} \\
\Sigma_{\mathbf{x}_i} &= \sigma_{x_k}^2 \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \\
\text{and } \sigma_{x_k}^2 &= Var(x_k) = Var(x_{k+1}) && (4.10)\\
\Rightarrow E[C] &= mP(x_{k+i} < L, x_{k+i+1} > L) \\
&= m \int_{-\infty}^{L} \int_{L}^{\infty} \mathcal{N}(\mathbf{x}_i; \mu_{\mathbf{x}_i}, \Sigma_{\mathbf{x}_i}) d\mathbf{x}_i && (4.11)
\end{aligned}
$$

Note that the most that Eqn. 4.9 can be simplified is shown in Eqn. 4.11. This result is not at all in closed form and requires numerical integration. Therefore, it would be very beneficial to apply our approximation for its continuous counterpart derived in the previous section. It can easy to see that Eqns. 4.12-4.14 now hold. The proof is exactly the same for the discretization of the derivative process. The only other additional part we need to complete the proof is stationarity, which is tacitly assumed in our use of Rice's formula. This provides us with the fact that $\mu_x = E[x(t)]$ in continuous time is the same in discrete time, $\mu_{x_k} = E[x_k]$, $\Rightarrow \mu_x = \mu_{x_k}$, and likewise $\sigma_x = \sigma_{x_k}$. Due to stationarity, these simple substitutions are good for the length of the interval in question. In other words, we are making a "truncated" zero-order hold (ZOH) approximation assumption over the time interval. An explanation of the "truncated" qualifier will follow shortly.

$$E[C] = \frac{mT_s}{2\pi}\frac{\sigma_{\dot{x}}}{\sigma_x}e^{-\frac{1}{2}\frac{(L-\mu_x)^2}{\sigma_x^2}} \tag{4.12}$$

$$\approx \frac{mT_s}{2\pi}\sqrt{\frac{2(1-\rho_1)}{T_s^2}}e^{-\frac{1}{2}\frac{(L-\mu_{x_k})^2}{\sigma_{x_k}^2}} \tag{4.13}$$

$$= \frac{m\sqrt{2(1-\rho_1)}}{2\pi}e^{-\frac{1}{2}\frac{(L-\mu_{x_k})^2}{\sigma_{x_k}^2}} \tag{4.14}$$

A graphic similar to Fig. 4.2 for this case cannot be illustrated without introducing a specific example. We can certainly show $E[C]$ as a function of μ_{x_k}, for different values of ρ_1. However, it is much easier to summarize the reduction of error in the approximation as a function of the one-step autocorrelation, ρ_1, as in Fig. 4.2. Therefore, we will save such illustrations for later, when we introduce a relevant example. The approximation shown in Eqn. 4.14 is very useful when developing discrete-time cost functions for optimal control. This approximation may be useful in our optimal control chapter, by assuming that the processes of interest are stationary and Gaussian.

4.2.3 Level Upcrossings of a Non-Stationary Gaussian Random Sequence

Finally, we must consider the most general case of a non-stationary Gaussian random sequence. This is necessary because the predicted process of a stationary Gaussian random sequence is actually non-stationary. Recall that predictions are necessary when dealing with optimal alarm design. Furthermore, the conditional predicted process, $x_{k+i|k}$, is not stationary because the variance can be shown to change with the length of the prediction window, i. Let's assume the same formulation as in the previous chapter, such that we will condition on all available information to obtain our prediction, via the Kalman filtering paradigm. Hence Eqns. 4.15-4.19 represent the corresponding exact discrete formula for the expected number of upcrossings of a non-stationary Gaussian random sequence.

$$
\begin{aligned}
E[C|x_0,\ldots,x_k] &= \sum_{i=0}^{m-1} P(x_{k+i} < L, x_{k+i+1} > L | x_0,\ldots,x_k) && (4.15)\\
&= \sum_{i=0}^{m-1} E[C_i|x_0,\ldots,x_k] && (4.16)\\
&= \sum_{i=0}^{m-1} g(\hat{x}_{k+i|k}, \hat{x}_{k+i+1|k}) && (4.17)\\
&= \sum_{i=0}^{m-1} \int_{-\infty}^{L}\int_{L}^{\infty} \mathcal{N}(\mathbf{x}_i; \mu_{\mathbf{x}_i}, \Sigma_{\mathbf{x}_i}) d\mathbf{x}_i && (4.18)\\
\text{where } \mathbf{x}_i &= \begin{bmatrix} x_{k+i|k} \\ x_{k+i+1|k} \end{bmatrix} \triangleq \begin{bmatrix} x_{k+i}|x_0,\ldots,x_k \\ x_{k+i+1}|x_0,\ldots,x_k \end{bmatrix} \in \mathbb{R}^2 \\
\mu_{\mathbf{x}_i} &= \begin{bmatrix} \hat{x}_{k+i|k} \\ \hat{x}_{k+i+1|k} \end{bmatrix} \\
\Sigma_{\mathbf{x}_i} &= \begin{bmatrix} Var(x_{k+i|k}) & cov(x_{k+i|k}, x_{k+i+1|k}) \\ cov(x_{k+i+1|k}, x_{k+i|k}) & Var(x_{k+i+1|k}) \end{bmatrix} && (4.19)
\end{aligned}
$$

This formula is very intuitively appealing, because we see how the expected number of upcrossings is related to the sum of the probabilities of upcrossings in each consecutive, adjacent pair of time slices in the entire prediction window. Unlike the stationary case, there is no further simplification when spanning m multiple time intervals. Furthermore, there is still no closed-form solution, and even when applying a similar approximation as before, there is no actual closed-form solution. However, we may develop a clever combination of approximations that reduces the computational burden significantly. The same idea of discretization is used, but we no longer can use the ZOH (zero-order hold) assumption. Rather, it will be much more useful to use a first-order hold (FOH) approximation due to the non-stationarity of the process, as well as another basic approximation that we'll introduce shortly.

Let's first present the formula for the expected number of upcrossings in an interval of a non-stationary Gaussian process in continuous time. This formula was first introduced in the famous Cramér and Leadbetter monograph, [14], on pg. 289, shown here as Eqn. 4.20. In the following

equations, $x(t)$ represents the nonstationary process of interest in continuous time, and L is the critical level. It is also known as the Leadbetter-Cryer integral.

$$
\begin{aligned}
E[C] &= \int_0^{T_s} \frac{r'(t)}{\eta(t)} \phi\left(r(t)\right) \left[\phi(\eta(t)) + \eta(t)\Phi(\eta(t))\right] dt \qquad (4.20)\\
\text{where } r(t) &= \frac{m(t)}{\sigma(t)}\\
\text{and } \eta(t) &= \frac{m'(t) - \gamma(t)\varrho(t)r(t)}{\gamma(t)\sqrt{1-\varrho^2(t)}}\\
m(t) &= E[x(t) - L]\\
\sigma^2(t) &= Var(x(t) - L)\\
\gamma^2(t) &= Var(\dot{x}(t))\\
\varrho(t) &= cor(x(t) - L, \dot{x}(t)) = \frac{cov(x(t) - L, \dot{x}(t))}{\sigma(t)\gamma(t)}\\
\phi(z) &= \text{Standard Normal PDF} = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{z^2}{2}}\\
\Phi(z) &= \text{Standard Normal CDF} = \int_{-\infty}^{z} \phi(\lambda) d\lambda
\end{aligned}
$$

Now let's perform the FOH discretization procedure based upon the interpolation formula shown in Eqn. 4.21. In the derivation we compute the expected number of level-crossings over a single time interval, such that $m = 1$, without loss of generality.

$$
\begin{aligned}
x(t) &\approx x_{k|k} + \frac{t}{T_s}(x_{k+1|k} - x_{k|k}) && (4.21)\\
\mathbf{b}^T &\triangleq \left[\begin{array}{cc} 1 - \frac{t}{T_s} & \frac{t}{T_s} \end{array}\right] && (4.22)\\
\Rightarrow m(t) &\approx \hat{x}_{k|k} - L + \frac{t}{T_s}(\hat{x}_{k+1|k} - \hat{x}_{k|k}) && (4.23)\\
&= \mathbf{b}^T \mu_{\mathbf{x}_i} - L && (4.24)\\
\Rightarrow m'(t) &\approx \frac{\hat{x}_{k+1|k} - \hat{x}_{k|k}}{T_s} && (4.25)\\
&= \frac{\partial \mathbf{b}^T}{\partial t}\mu_{\mathbf{x}_i} && (4.26)\\
\sigma^2(t) &\approx \mathbf{b}^T \Sigma_{\mathbf{x}_i} \mathbf{b} && (4.27)\\
\gamma^2(t) &\approx \frac{\partial \mathbf{b}^T}{\partial t}\Sigma_{\mathbf{x}_i}\frac{\partial \mathbf{b}}{\partial t} && (4.28)\\
\sigma(t)\sigma'(t) &\approx \frac{\partial \mathbf{b}^T}{\partial t}\Sigma_{\mathbf{x}_i}\mathbf{b} = \mathbf{b}^T\Sigma_{\mathbf{x}_i}\frac{\partial \mathbf{b}}{\partial t} && (4.29)
\end{aligned}
$$

$$
\begin{aligned}
\varrho(t) &\approx \frac{\sigma'(t)}{\gamma(t)} = \frac{\frac{\partial \mathbf{b}^T}{\partial t}\Sigma_{\mathbf{x}_i}\mathbf{b}}{\sqrt{\mathbf{b}^T\Sigma_{\mathbf{x}_i}\mathbf{b}}} && (4.30)\\
r(t) &\approx \frac{\mathbf{b}^T\mu_{\mathbf{x}_i} - L}{\sqrt{\mathbf{b}^T\Sigma_{\mathbf{x}_i}\mathbf{b}}} && (4.31)
\end{aligned}
$$

Now that we've approximated all of the terms of the Leadbetter-Cryer integral in discrete-time by performing a FOH discretization, we need to invoke a clever way to simplify computing this integral.

$$
\begin{aligned}
\text{Let } g(t) &\triangleq \frac{r'(t)}{\eta(t)}\left[\phi(\eta(t)) + \eta(t)\Phi(\eta(t))\right] && (4.32)\\
\Rightarrow E[C] &= \int_0^{T_s} \phi(r(t))g(t)dt && (4.33)\\
\text{Recall } r(t) &= \frac{m(t)}{\sigma(t)} && (4.34)
\end{aligned}
$$

Let's now make a convenient simultaneous definition of $r(t)$, where we tacitly now redefine t as a Gaussian random variable.

$$\text{Let } r(t) = \frac{t-\mu}{s} = \frac{m(t)}{\sigma(t)} \tag{4.35}$$

$$\text{where } \mu = t^* : m(t^*) = 0 \text{ (Solve with Eqns. 4.23-4.24)} \tag{4.36}$$

$$\Rightarrow s = \lim_{t\to\mu} \frac{t-\mu}{r(t)} \tag{4.37}$$

$$= \frac{1}{r'(\mu)} \tag{4.38}$$

$$\Rightarrow E[C] = \int_0^{T_s} \phi(\frac{t-\mu}{s}) g(t) dt \tag{4.39}$$

$$= \int_0^{T_s} \underbrace{\frac{1}{s}\phi(\frac{t-\mu}{s})}_{\mathcal{N}(t;\mu,s^2)} \overbrace{sg(t)}^{f(t)\triangleq} dt \tag{4.40}$$

$$= \int_0^{T_s} \mathcal{N}(t;\mu,s^2) f(t) dt \tag{4.41}$$

$$= \left(\int_0^{T_s} \mathcal{N}(t;\mu,s^2) dt\right) (E[f(t)|0<t<T_s]) \tag{4.42}$$

$$\approx \left[\Phi\left(\frac{T_s-\mu}{s}\right) - \Phi\left(\frac{-\mu}{s}\right)\right] f(\mu) \tag{4.43}$$

The final approximation was made by using the fact that $f(\mu) \approx E[f(t)|0<t<T_s]$ if $f(t)$ is sufficiently smooth over the region of integration [9, 56].

$$\Rightarrow E[C] \approx \left[\Phi\left(\frac{T_s-\mu}{s}\right) - \Phi\left(\frac{-\mu}{s}\right)\right] sg(\mu) \tag{4.44}$$

$$= \left[\Phi\left(\frac{T_s-\mu}{s}\right) - \Phi\left(\frac{-\mu}{s}\right)\right] \frac{sr'(\mu)}{\eta(\mu)} \left[\phi(\eta(\mu)) + \eta(\mu)\Phi(\eta(\mu))\right] \tag{4.45}$$

$$\text{Recall } s = \frac{1}{r'(\mu)} \tag{4.46}$$

$$\Rightarrow E[C] \approx \left[\Phi\left(\frac{T_s-\mu}{s}\right) - \Phi\left(-\frac{\mu}{s}\right)\right] \left[\frac{\phi(\eta(\mu))}{\eta(\mu)} + \Phi(\eta(\mu))\right] \tag{4.47}$$

The final expression is simply a nonlinear function of s and μ. Although not quite in closed form due to the presence of $\Phi(\cdot)$, there is an extreme computational advantage by solving this equation in lieu of the integral in its original form. Standard routines to compute $\Phi(\cdot)$ via the error function, **erf**$(\cdot)$ are pervasive and very computationally efficient. Recall how μ is found: $\mu = t^* : m(t^*) = 0$. The expression was found by using the fact that the mean of a Gaussian can

be found at its maximum, so we can take the derivative of $\phi(r(t))$ and set it to zero. This gives $-r(t^*)\phi(r(t^*))r'(t^*) = 0 \Rightarrow r(t^*) = 0 \Rightarrow m(t^*) = 0$. Therefore, we can easily find the location of the maximum, and hence the mean is $\mu = t^*$, where $t \sim \mathcal{N}(\mu, s^2) \Rightarrow \phi(r(t)) = \phi(\frac{t-\mu}{s})$. This is solved with Eqns. 4.23-4.24, giving $\mu = \frac{T_s(L-\hat{x}_{k|k})}{\hat{x}_{k+1|k}-\hat{x}_{k|k}}$. Furthermore, $s = \frac{1}{r'(\mu)} = \left[\frac{\sigma'(\mu)}{\sigma(\mu)}\left(\frac{m'(\mu)}{\sigma'(\mu)} - \frac{m(\mu)}{\sigma(\mu)}\right)\right]^{-1}$. Hence, $E[C]$ is a multivariate function of $\hat{x}_{k+1|k}$ and $\hat{x}_{k|k}$, as we'd expect.

4.3 Examples

4.3.1 Zero-Level Upcrossings of a Zero-Mean Stationary Gaussian Random Sequence

For the expected number of zero-level upcrossings of a zero-mean stationary Gaussian random sequence, there is no specific example required in order to determine the quantitative or qualitative nature of approximation error. This is evident by observing Fig. 4.2, shown again in Fig. 4.3 for convenience. Table 4.1 shows the quantification of approximation error at specific values of ρ_1, where $f(\rho_1) = \frac{1}{4} - \frac{1}{2\pi}\arcsin(\rho_1)$, and $\hat{f}(\rho_1) = \frac{\sqrt{2(1-\rho_1)}}{2\pi}$.

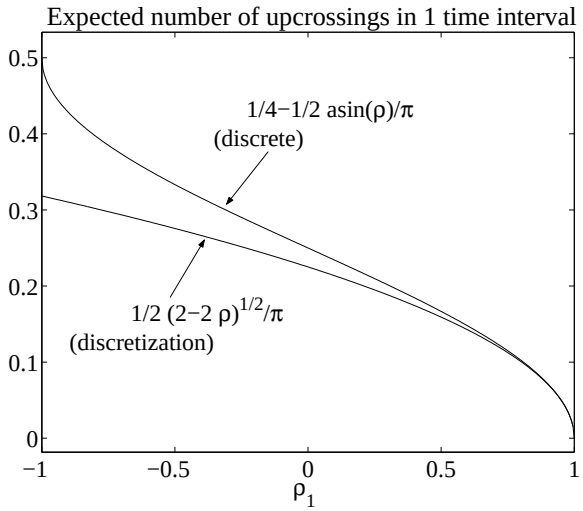

Figure 4.3: Discretization vs. Exact Discrete Level Upcrossings (Courtesy Repeat)

ρ_1	$\lvert f(\rho_1) - \hat{f}(\rho_1)\rvert$
-1	0.1817
-0.5	0.0577
0	0.0249
0.5	0.0075
0.75	0.0025
0.9	6.0689×10^{-4}
0.95	2.121×10^{-4}
0.99	1.88×10^{-5}
0.999	5.9327×10^{-7}

Table 4.1: Table of Approximation Error for $E[C]$

We therefore see that for this simple case, the approximation error is negligible for systems that have the property of $\rho_1 \gtrapprox 0.9$.

4.3.2 Level Upcrossings of a Stationary Gaussian Random Sequence

For the case in which computing the expected number of non-zero level upcrossings of a non-zero mean stationary Gaussian random sequence is desired, a specific example is required to demonstrate the quantitative and qualitative nature of approximation error. The example to be used in studying the approximation error characteristics of the expected number of level-crossings for the non-zero level and non-zero mean case was discussed in Sec. 3.3.2. Recall the measurement noise was modelled in this section, and that the only reason we were concerned with it was to perform Kalman filtering and prediction. Furthermore, the original continuous time model and level-crossing formulae used by Federspiel [19, 22] do not account for measurement noise. Therefore, let us disregard the measurement noise in our current example.

It is instructive to provide an illustration as to how the expected number of complaints varies with the autocorrelation, as we saw with the simple case before in Fig. 4.3. Let's choose the same sampling interval, $T_s = 20$ min, as was used in Sec. 3.3.2 as a fundamental parameter. Varying the autocorrelation of the differences between the complaint and building processes artificially then yields Fig. 4.4. This illustrates the sum of the expected number of cold and hot complaints at

the two peak values (mean of T_H and T_L) as a function of ρ_h. Here, T_H and T_L are the hot and cold complaint level processes, respectively. Also, ρ_h is the one-step autocorrelation of the process difference, and $z_H = T_B - T_H$, where T_B is the building temperature process.

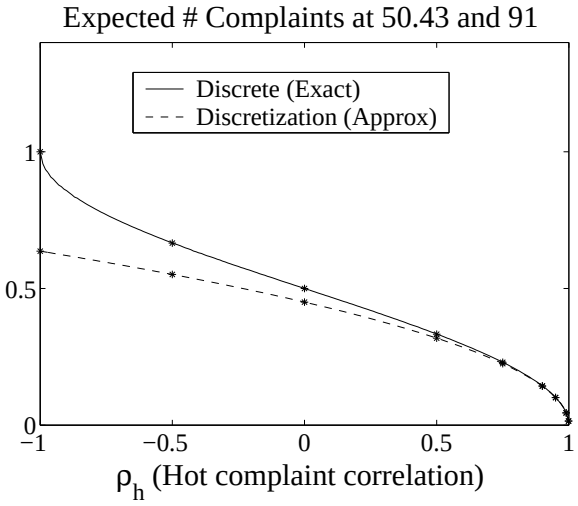

Figure 4.4: Peak Expected Complaints vs. hot correlation, ρ_h

In Fig. 4.4 we see how the approximation improves with increasing ρ_h, just as before in Fig. 4.3. However, it may be more intuitive and useful to get an idea of what happens to the approximations as we vary the sampling interval over a range of values in lieu of artificially varying ρ_h. It is also important to demonstrate how well both discrete versions of the cost function approximate the continuous cost function with decreased sampling interval size. The continuous complaint level and building temperature dynamics are truly what the original model was meant to characterize.

As the sampling interval decreases, the continuous cost function acts a limiting function for both discrete cost functions. This is evident in Fig. 4.5, where several cost functions are shown explicitly for different values of ρ_h. The cost functions illustrated in this figure are shown for a fixed complaint frequency. Alternatively, Figs. 4.6 and 4.7 show the resulting hot and cold cost function values evaluated at peaks on the bottom two graphs of each figure, expressed as a function of the

autocorrelation ρ_h and ρ_l, respectively. Fig. 4.7 is based on the same data shown in Fig. 4.6, but the range of the y-axis for the bottom two graphs is limited in Fig. 4.7, to illustrate more detail. It is evident that when $\rho_1 \to 1$, both discrete versions of the peak cost function approach the continuous result. In fact, we can also see that the bottom two graphs mirror the same characteristic as in Figs. 4.4 and 4.3 of the discretization approximation approaching the exact discrete result as $\rho_1 \to 1$.

It is clear from the data labels provided in the bottom two graphs of Figs. 4.6 and 4.7 that decreasing the sampling interval has the same effect as increasing the autocorrelation. For convenience, we've also shown the autocorrelation as a function of the sampling interval on the top two graphs in both figures, so we can get an idea of the monotonicity of $\rho_h(T_s)$ and $\rho_l(T_s)$. This also illustrates the inverse relationship between $\rho_h(T_s)$, $\rho_l(T_s)$ and the sampling interval, T_s.

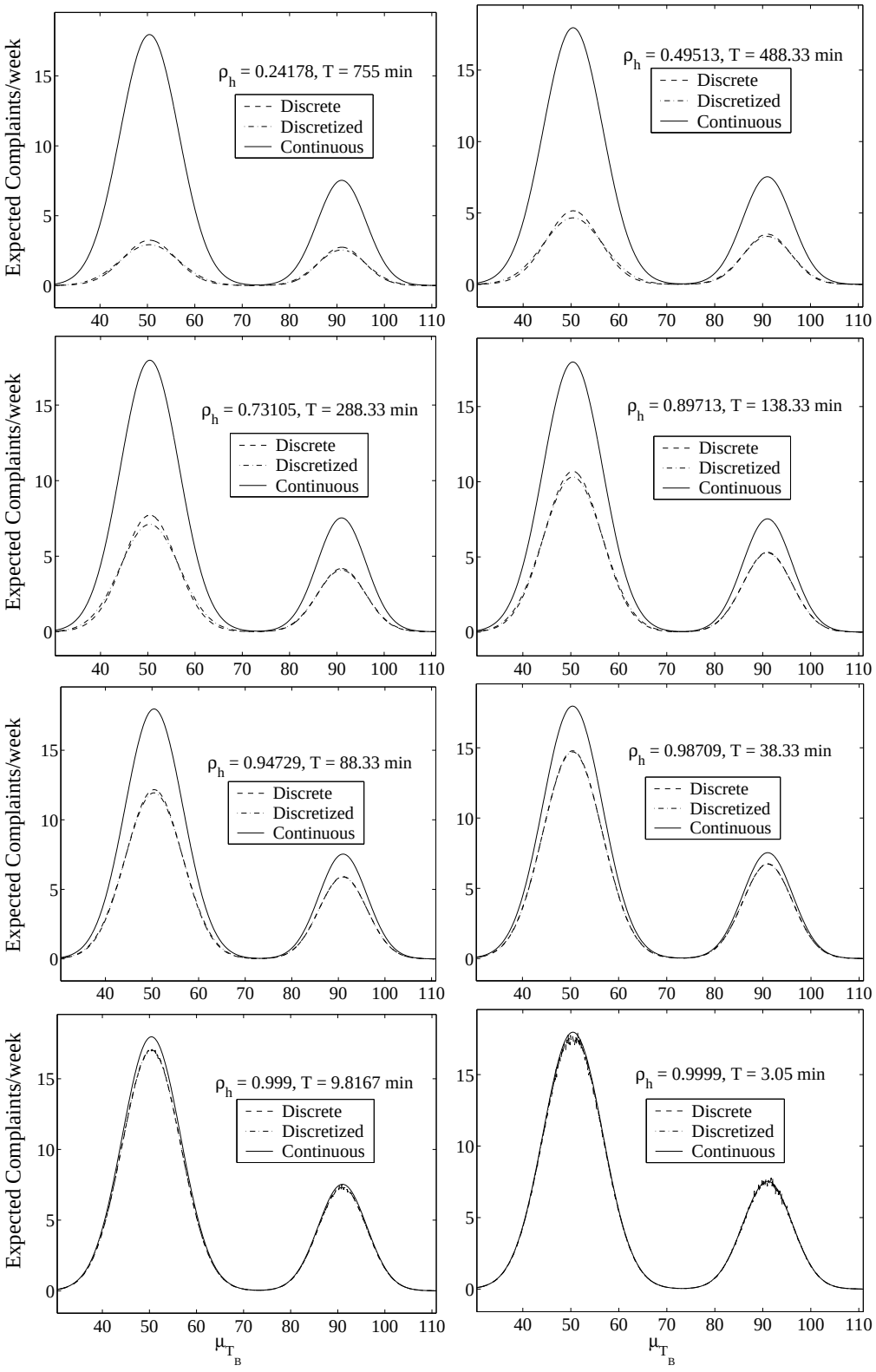

Figure 4.5: Cost Function Approximations for Expected Complaints/week vs. μ_{T_B}

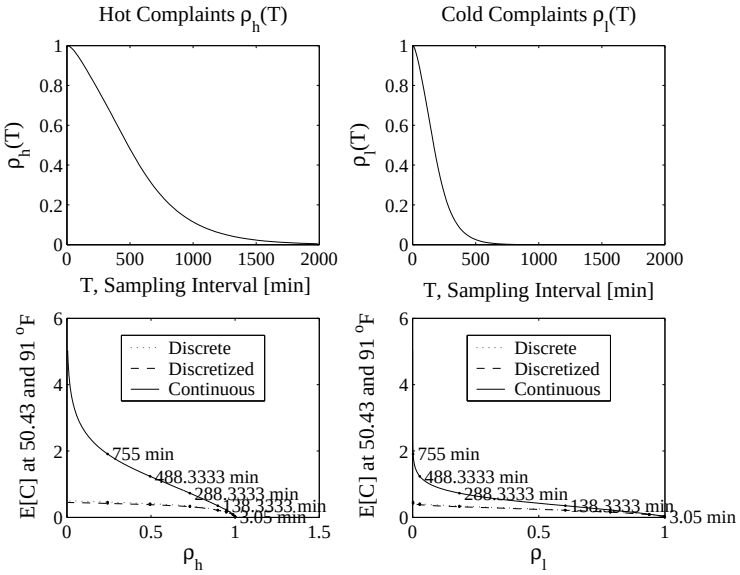

Figure 4.6: Approximation Summary: $\rho(T_s)$ and Peak $E[C]$ vs. ρ

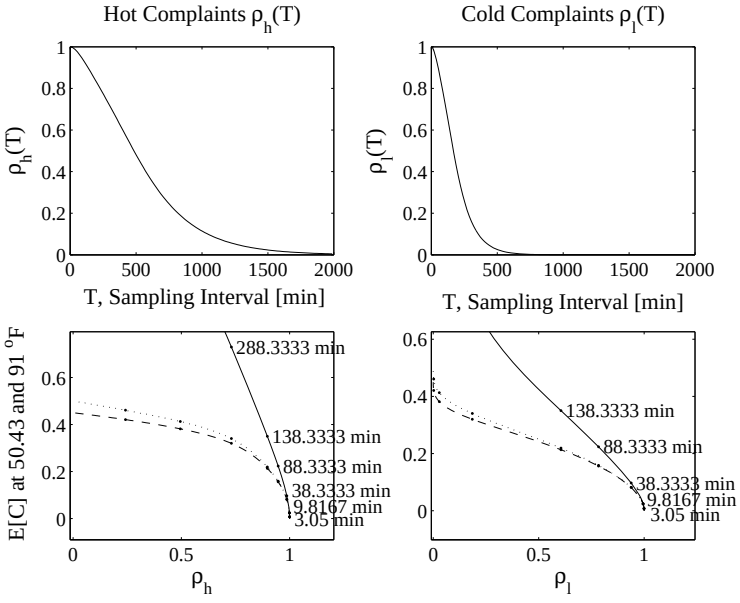

Figure 4.7: Approximation Summary Detail: $\rho(T_s)$ and Peak $E[C]$ vs. ρ

If the measurement noise had been modelled, there would have been an extreme bias

introduced for small sampling intervals due to reasons discussed in Sec. 3.3.2. Therefore, we shall only consider modelling measurement noise for implementation in future work when it is required for estimation of the state via Kalman filtering. When implemented, the sampling interval should be selected such that it is large enough to prevent biasing the results, but small enough to reasonably approximate the exact discrete and continuous-time formulae. The measurement noise will then maintain its "whiteness" due to the relative magnitude of the correlation time.

4.3.3 Level Upcrossings of a Non-Stationary Gaussian Random Sequence

Unlike the previous section, we are no longer concerned with the continuous formula for level upcrossings of a non-stationary Gaussian random sequence. The original model and proposed cost function developed by Federspiel [19, 22] made the assumption that the process of interest was stationary. The main goal of the model was for use in prediction and control. Here, the discretization of the formula for level upcrossings of a non-stationary Gaussian random sequence is mainly useful in the following chapter, where we present an alternative alarm design metric, $E[C|A]$.

Therefore, we will present and compare the results of level-crossings of a non-stationary Gaussian random sequence using the exact discrete formula and the discretization approximation. Recall from Sec. 4.2.3 that a predicted process of a stationary Gaussian random sequence is actually non-stationary, and predictions are necessary when dealing with optimal alarm design. Furthermore, the conditional predicted process, $x_{k+i|k}$, is not stationary because the variance changes with the length of the prediction window, i. Therefore, we use the same formulation as in the previous chapter, by conditioning on all available information to obtain our prediction, via the Kalman filtering paradigm. Recall Fig. 3.1 from the previous chapter, shown again in Fig. 4.8 for convenience.

Contour curves similar to the ones shown in Fig. 4.8 will act as the baseline example for the expected number of level upcrossings of a non-stationary Gaussian random sequence. The contour curves show only the heights of the actual number of level upcrossings. They must be computed numerically since they are based upon the exact discrete formulae for the number of level-upcrossings

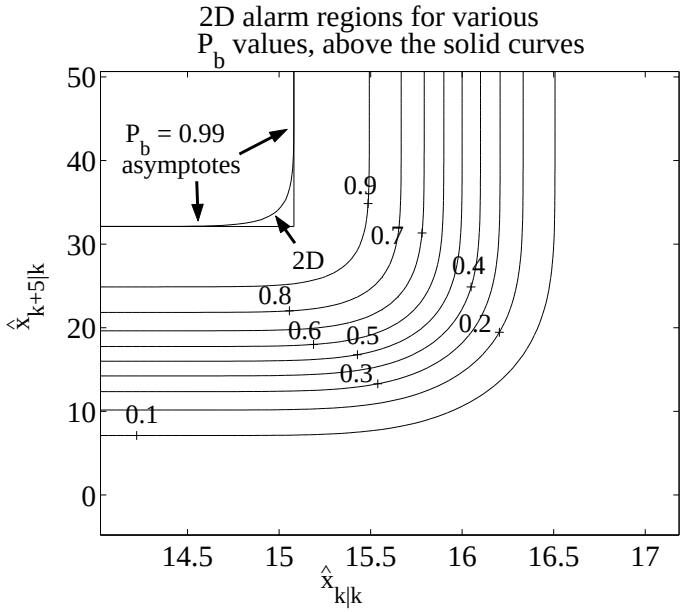

Figure 4.8: Approximation to the 2D Alarm Region for $P_b = 0.99$

in a single time interval. Eqns. 4.48-4.50 characterize the single time interval formula necessary for this computation, which we use without loss of generality.

$$g(\hat{x}_{k+i}, \hat{x}_{k+i+1}) = E[C_i | x_0, \ldots, x_k] \tag{4.48}$$

$$= P(x_{k+i} < L, x_{k+i+1} > L | x_0, \ldots, x_k) \tag{4.49}$$

$$= \int_{-\infty}^{L} \int_{L}^{\infty} \mathcal{N}(\mathbf{x}_i; \mu_{\mathbf{x}_i}, \Sigma_{\mathbf{x}_i}) d\mathbf{x}_i \tag{4.50}$$

$$\text{where } \mathbf{x}_i = \begin{bmatrix} x_{k+i|k} \\ x_{k+i+1|k} \end{bmatrix} \triangleq \begin{bmatrix} x_{k+i} | x_0, \ldots, x_k \\ x_{k+i+1} | x_0, \ldots, x_k \end{bmatrix} \in \mathbb{R}^2$$

$$\mu_{\mathbf{x}_i} = \begin{bmatrix} \hat{x}_{k+i|k} \\ \hat{x}_{k+i+1|k} \end{bmatrix}$$

$$\Sigma_{\mathbf{x}_i} = \begin{bmatrix} Var(x_{k+i|k}) & cov(x_{k+i|k}, x_{k+i+1|k}) \\ cov(x_{k+i+1|k}, x_{k+i|k}) & Var(x_{k+i+1|k}) \end{bmatrix}$$

By using the semi-parametric expression derived in the previous section, shown below as Eqns. 4.51-4.53, we can come up with a FOH approximation of the exact discrete contour curves based on Eqns. 4.48-4.50. The equation is considered to be "semi-parametric" due to the presence of the $\Phi(\cdot)$ normal CDF.

$$E[C] \approx \left[\Phi\left(\frac{T_s-\mu}{s}\right)-\Phi\left(-\frac{\mu}{s}\right)\right]\left[\frac{\phi(\eta(\mu))}{\eta(\mu)}+\Phi(\eta(\mu))\right] \tag{4.51}$$

$$\text{where } \mu = \frac{T_s(L-\hat{x}_{k|k})}{\hat{x}_{k+1|k}-\hat{x}_{k|k}} \tag{4.52}$$

$$\text{and } s = \frac{1}{r'(\mu)} = \left[\frac{\sigma'(\mu)}{\sigma(\mu)}\left(\frac{m'(\mu)}{\sigma'(\mu)}-\frac{m(\mu)}{\sigma(\mu)}\right)\right]^{-1} \tag{4.53}$$

Hence, the semi-parametric approximation for $E[C]$ is a multivariate function of $\hat{x}_{k+1|k}$ and $\hat{x}_{k|k}$, similar to the exact discrete formula for $E[C_i|x_0,\ldots,x_k] = g(\hat{x}_{k+i},\hat{x}_{k+i+1})$ as we'd expect. The computational effort required to solve this approximation is much less than required for the numerical computations. The approximations will be very useful when coming up with a new alarm design metric to presented in the subsequent chapter. In fact, when given the correct circumstances of $\rho_1 \approx 1$, the contour curves may also be used to approximate the 2-dimensional alarm regions in lieu of using the "asymptotes" as presented before. This would have an impact on any approximation scheme presented that relied on using these asymptotes. Specifically, two approximations in the class of optimal alarm systems would benefit from this finding: the 2-dimensional alarm system and the multiple sub-interval approximation. However, recall from Chap. 3 that using this approximation to perform an integration over the 2D alarm region will still have a greater computational burden than using asymptotes. Therefore, all of the examples presented in Chap. 3, and the results to be discussed in Chap. 7 are based upon the latter method.

As before, it is instructive to provide an illustration of the expected number of complaints vs. the autocorrelation. Again, it is more intuitive and useful to vary the sampling interval over a range of values in lieu of artificially varying ρ_h. Recall that decreasing the sampling interval has the same effect as increasing the autocorrelation. Fig. 4.9 illustrates the contour lines of several

cost functions shown explicitly for different values of ρ_h and corresponding sampling interval, T_s. The contour plots of the cost functions of interest are now only showing the expected number of hot complaints as opposed to both hot and cold, to simplify the curves in the graph. In the graphs, the dotted lines represent the exact discrete numerical computations of the number of level-upcrossings in a single time interval. The solid lines represent their first-order hold (FOH) discretization approximations for the given sampling intervals and corresponding autocorrelations shown.

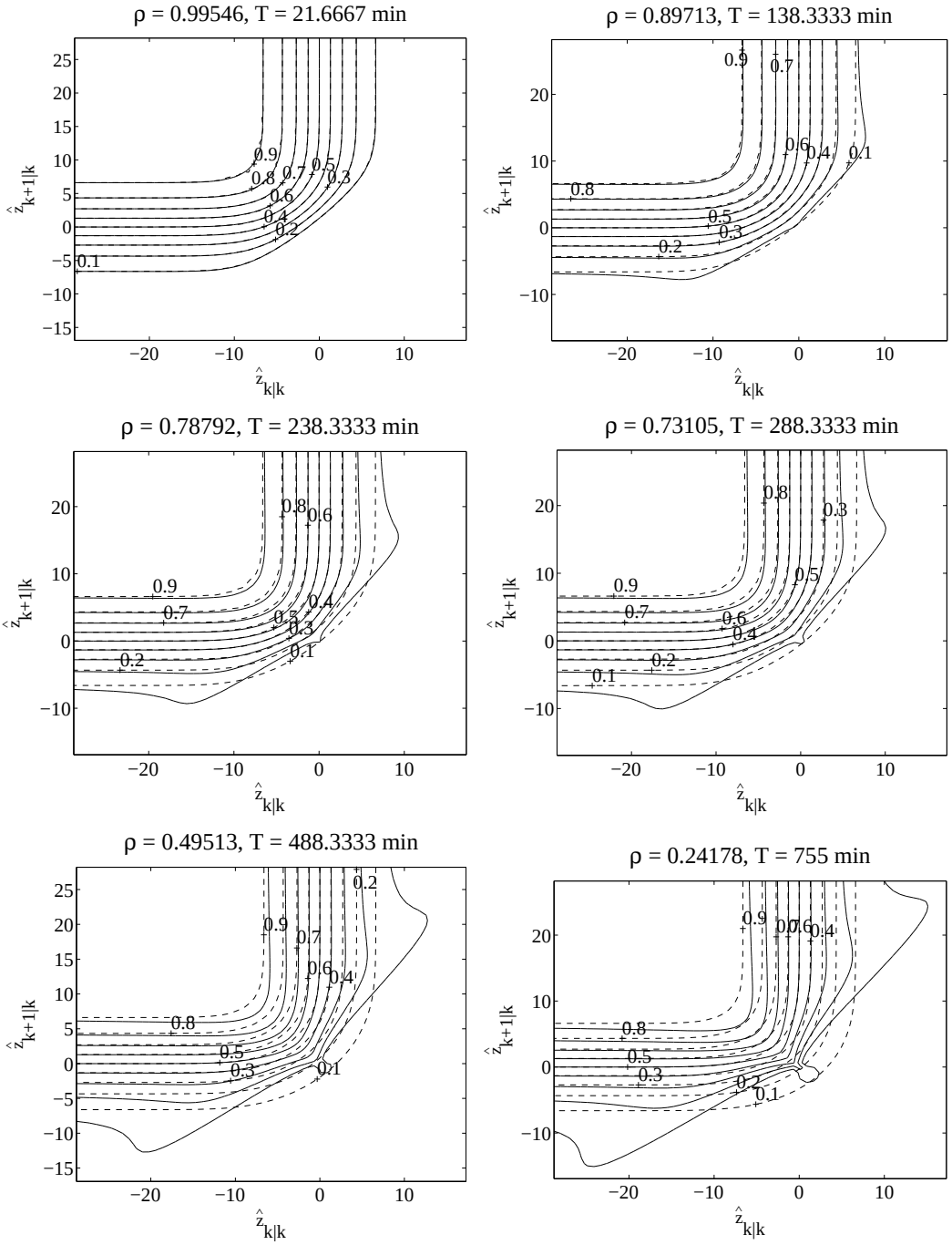

Figure 4.9: Contour Plots of FOH Approximations for Non-Stationary Expected Complaints

As seen in Fig. 4.9, any sampling interval of $T_s \approx 20$ min or less gives a sufficiently good

FOH approximation of the exact discrete result for computing the expected number of complaints of a non-stationary Gaussian random sequence. Recall that this is the sampling interval used in our example from Sec. 3.3.2. As such, the dynamics and sampling interval parameter, $T_s = 20$ min of interest merit the use of this approximation for the optimal alarm design metric to be used in the subsequent chapter.

4.4 Discussion

It is important to note that the discretizations introduced in this chapter are truly "mixed." This statement is meant to address the fact that the model presented in Sec. 3.3.2 actually uses a true "sampled-data" zero-order hold discretization procedure of the continuous time model. This discretization can be shown to be much more accurate than the very straightforward and simple approximation of Eqn. 4.8. The latter can be considered a "truncated" discretization of the derivative, used to arrive at our discretized cost function. A more detailed explanation of these two different procedures is provided below, and it will become clear why we've worded them as such. It will also be shown how they are related, but not really the same. The so-called "sampled-data" ZOH discretization of a state-space model was performed in Sec. 3.3.2. Recall the results shown below again for convenience.

$$\mathbf{z}_{k+1} = \mathbf{A}_d \mathbf{z}_k + \mathbf{B}_d n_k \tag{4.54}$$

$$y_k = \mathbf{C}_d \mathbf{z}_k \tag{4.55}$$

$$\mathbf{q}_{k+1} = \mathbf{A}_d \mathbf{q}_k + \mathbf{B}_d u_k + w_k \tag{4.56}$$

$$x_k = \mathbf{C}_d \mathbf{q}_k + v_k \tag{4.57}$$

where

$$\mathbf{A}_d = e^{\mathbf{A}\tau_s} \tag{4.58}$$

$$\mathbf{B}_d = (e^{\mathbf{A}\tau_s} - \mathbf{I})\mathbf{A}^{-1}\mathbf{B} \tag{4.59}$$

$$\mathbf{C}_d = \mathbf{C} \tag{4.60}$$

Also recall the formula for the matrix exponential is simply a Taylor series expansion.

$$e^{\mathbf{A}\tau_s} = \mathbf{I} + \tau_s\mathbf{A} + \frac{(\tau_s\mathbf{A})^2}{2!} + \frac{(\tau_s\mathbf{A})^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{(\tau_s\mathbf{A})^n}{n!}$$

Note the use of the matrix exponential in defining $\mathbf{A}_d$ and $\mathbf{B}_d$ in Eqns. 4.58 and 4.59 above. If we truncate the Taylor series using a first-order truncation, we then obtain a poor approximation to the matrix exponential.

$$e^{\mathbf{A}\tau_s} \approx \mathbf{I} + \tau_s\mathbf{A}$$

Now let's see what happens when we find a discretized state-space system based upon the straightforward and simple approximation of Eqn. 4.8. Let's first start with the continuous-time systems, as shown in Eqns. 4.61-4.64 below.

$$\dot{\mathbf{z}}(t) = \mathbf{A}\mathbf{z}(t) + \mathbf{B}n(t) \tag{4.61}$$

$$y(t) = \mathbf{C}\mathbf{z}(t) \tag{4.62}$$

$$\dot{\mathbf{q}}(t) = \mathbf{A}\mathbf{q}(t) + \mathbf{B}u(t) + \mathbf{B}_w w(t) \tag{4.63}$$

$$x(t) = \mathbf{C}\mathbf{q}(t) + v(t) \tag{4.64}$$

By using Eqn. 4.8, $\dot{\mathbf{z}}(t) \approx \frac{\mathbf{z}_{k+1}-\mathbf{z}_k}{T_s}$, and $\dot{\mathbf{q}}(t) \approx \frac{\mathbf{q}_{k+1}-\mathbf{q}_k}{T_s}$, we can easily find an alternate discretized state-space system. It can be found by using the simple approximation to the first derivative above, and is clearly equivalent to the first-order Taylor series truncation in the following equations.

$$
\begin{aligned}
\mathbf{z}_{k+1} &= \underbrace{(\mathbf{I}+\tau_s\mathbf{A})}_{\mathbf{A}_d}\mathbf{z}_k + \underbrace{\tau_s\mathbf{B}}_{\mathbf{B}_d} n_k \\
y_k &= \underbrace{\mathbf{C}}_{\mathbf{C}_d}\mathbf{z}_k \\
\mathbf{q}_{k+1} &= \underbrace{(\mathbf{I}+\tau_s\mathbf{A})}_{\mathbf{A}_d}\mathbf{q}_k + \underbrace{\tau_s\mathbf{B}}_{\mathbf{B}_d} u_k + w_k \\
x_k &= \underbrace{\mathbf{C}}_{\mathbf{C}_d}\mathbf{q}_k + v_k
\end{aligned}
$$

Note that $e^{\mathbf{A}\tau_s} \approx \mathbf{I} + \tau_s\mathbf{A} = \mathbf{A}_d$ in the state-space equations above. Therefore, we can see the direct relationship between the truncated discretization and the sampled data discretization via the number of terms used in the Taylor series expansion of the matrix exponential. In fact, if we used all of the terms of the matrix exponential, our approximation to the derivative would be exact, then accounting for all higher order terms. However, this truncated state-space discretization is *not* the one we actually use. The truncated discretization is performed only for the approximation to the derivative used in deriving the cost function. Hence, this explains why in reality we have a "mixed discretization." The sampled data ZOH discretization has to do with the continuous dynamics of the state-space model, and the truncated discretization serves the purpose of approximating the cost function. A FOH discretization is used in lieu of the truncated ZOH discretization when applied to the cost function characterizing level-crossing of a non-stationary Gaussian process. This approximation is used in practice for a new optimal alarm design metric.

4.5 Conclusion

This chapter has introduced some new approximations that establish useful relationships between discrete-time optimal control and alarm theory and continuous time-level crossing formulae. Such approximations may not have been considered in the past because there currently exist discrete-time level-crossing formula that provide exact results. However, in specific cases using these formulae are computationally burdensome and therefore not very practical, especially for the imple-

mentation of real-time optimal control algorithms and optimal alarm design methods. Therefore, this chapter has introduced some very straightforward approximations that provide an alternative to the exact discrete-time level crossing formulae that are computationally much more favorable. All approximations introduced are best for systems that have an autocorrelation close to 1. This can be interpreted as any system/process that has a high data sampling rate, or a small sampling interval.

Some other contributions include the fact that the non-stationary approximation can be used to parameterize the alarm regions in optimal alarm design. Previously thought to be unparameterizable [69], when the sampling rate is sufficiently high, this approximation serves as excellent semi-parametric representation of the alarm region. In this case, no other hyperbolic or rectangular approximations to the alarm region are required, as were used in the previous chapter. Furthermore, we'd like to be able to see the practical use of the FOH approximation in optimal alarm design. In the next chapter we introduce a new cost function to use in designing an optimal alarm system when one's primary concern is the minimization of false alarms.

Chapter 5

Optimal Alarm Design

5.1 Introduction

Some practical issues need to be considered when an optimal alarm system is being designed for a real engineering problem. When computing the probabilities of alarm, false alarm, and missed detection ($P(A)$, $P(C'|A)$, and $P(A'|C)$ respectively) the prediction window often includes a number of steps. Because of this, there are two main issues that present themselves. As explained in Chap. 3, accurate computation of these probabilities numerically require the inclusion/exclusion rule. Therefore, the number of terms grows exponentially with the number of steps in the prediction window, $\mathcal{O}(2^m)$, where m = the number of steps in the prediction window. This is only the case if we are using the multi-dimensional method of computing such probabilities, as explained in Chap. 3. If we use the sub-interval method, then the number of terms needed grows as $\mathcal{O}(2^{N_s})$, where N_s = the number of sub-intervals in the prediction window. These two methods are the most accurate methods to use in computing approximations of the exact probabilities in question. Therefore, we are faced with the curse of dimensionality, and potentially have a very long design time for a modest-sized prediction window !

Furthermore, these probabilities also need to be computed with high accuracy, therefore the

integration method used to compute them [27] requires a large number of sample points, extending the design time even further. As discussed in Chap. 3, the probability of missing at least one level-crossing over a long prediction window is very high over a large portion of the range of possible P_b values. This is a byproduct of the fact that the analysis of detections and alarms are very sensitive to the border probability, such that the useful range of border probabilities that the probability of missed detection spans will be very small and short. All of this makes it very difficult to design a robust optimal alarm system, since it is based upon probabilities that are so numerically sensitive.

5.2 Design Metrics for the Optimal Alarm System

5.2.1 Baseline Design Methods

As stated earlier, the steps required for design of an alarm system have been covered in previous work [69, 70]. Suggestions for choosing an optimal value of P_b in the design of the alarm system were provided in this literature. However, they tend to be based either on purely heuristic trial-and-error approaches, or other cost functions. Alarm system metrics are defined as $P(C|A)$ (correct alarm) and $P(A|C)$ (correct detection), or their complements, $P(C'|A)$ (false alarm, or Type I error) and $P(A'|C)$ (missed detection, or Type II error).

Iterative Approach

The heuristic trial-and-error approach is implemented according to the following iterative procedure:

1. Select a desired value for the probability of correct detection, P_d.
2. Select a new value for the alarm design parameter, P_b.
3. Find the alarm region or its approximation, based on P_b.

4. Compute the corresponding relevant alarm system metrics (operating characteristic functions), making sure to find the value of $P(A|C)$ corresponding to the current value of P_b.

5. Check to see if $P(A|C) \geq P_d$.

6. If this inequality does not hold, return to step 2, and continue iterating until the condition is met.

Obviously we can do better than this trial-and-error approach. This was recognized by Svensson [69], and he gave an alternative approach, by using cost functions.

Cost Functions

For cost functions, sometimes it is easy to assign particular costs, or weights to events that penalize the probability of alarms, false alarms, and missed detections, etc. Let's make some basic definitions prior to introducing a typical cost function.

$$
\begin{aligned}
P(A) &= \text{Probability of an Alarm} \\
P(C'|A) &= \text{Probability of a False Alarm (Type I Error Probability)} \\
P(C) &= \text{Probability of Complaints} \\
P(A'|C) &= \text{Probability of a Missed Detection (Type II Error Probability)} \\
\gamma_A &= \text{Cost of an Alarm} \\
\gamma_{FA} &= \text{Cost of a False Alarm} \\
\gamma_C &= \text{Cost of Complaints} \\
\gamma_{MD} &= \text{Cost of a Missed Detection} \\
\text{Let } \gamma'_{FA} &= \gamma_{FA} P(A) \\
\text{Let } \gamma'_{MD} &= \gamma_{MD} P(A)
\end{aligned}
$$

We know that $P(A)$, $P(C'|A)$, and $P(A'|C)$, can be parameterized monotonically by P_b, as a set of "basis functions" in some sense. This is evident from Fig. 5.1. Therefore, we can express

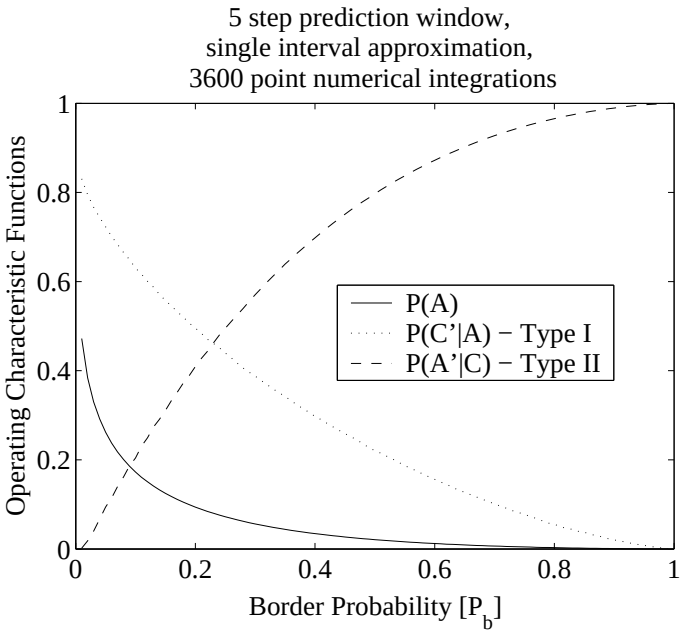

Figure 5.1: Parametrization of P_b for $P(A)$, $P(C'|A)$, and $P(A'|C)$

these alarm-based metrics as functions of P_b, as follows:

$$f_A(P_b) \triangleq P(A)$$

$$f_{FA}(P_b) \triangleq P(C'|A)$$

$$f_{MD}(P_b) \triangleq P(A'|C)$$

We'll also need to define the probability of complaints **and** alarm:

$$f_{CA}(P_b) \triangleq P(C, A)$$

A cost function can now be expressed as function of P_b.

$$\begin{aligned} J(P_b) &= \gamma_A f_A(P_b) + \gamma'_{FA} f_{FA}(P_b) + \gamma_C P(C) + \gamma'_{MD} f_{MD}(P_b) \quad (5.1) \\ &= (\gamma_A + \gamma_{FA}) f_A(P_b) - (\gamma_{FA} + \gamma_{MD}) f_{CA}(P_b) + (\gamma_C + \gamma_{MD}) P(C) \end{aligned}$$

The optimization problem is thus to find:

$$P_b^{opt} = \arg\min_{P_b} J(P_b)$$

Notice that Eqn. 5.1 contains a term, $(\gamma_C + \gamma_{MD})P(C)$, that is not a function of P_b, and therefore adds no utility to the optimization problem posed. Therefore, we can define $J'(P_b) \triangleq (\gamma_A + \gamma_{FA}) f_A(P_b) - (\gamma_{FA} + \gamma_{MD}) f_{CA}(P_b)$ and reformulate the optimization problem as follows:

$$\begin{aligned} P_b^{opt} &= \arg\min_{P_b} J(P_b) \quad (5.2) \\ &= \arg\min_{P_b} J'(P_b) \quad (5.3) \\ &= \arg\min_{P_b} \underbrace{(\gamma_A + \gamma_{FA})}_{\gamma_1} f_A(P_b) - \underbrace{(\gamma_{FA} + \gamma_{MD})}_{\gamma_2} f_{CA}(P_b) \quad (5.4) \\ &= \arg\min_{P_b} \gamma_1 f_A(P_b) - \gamma_2 f_{CA}(P_b) \quad (5.5) \end{aligned}$$

Hence, solving this constrained optimization problem over the values $P_b \in [0, 1]$ will lead to a tradeoff between the fewest possible alarms, and the greatest number of correct alarms, when posed as a minimization. An example of such a cost function is shown in Fig. 5.2, based upon the "basis functions" illustrated in Fig. 5.1. It is shown for several different cost weights of γ_1, and γ_2, where $\gamma_1 + \gamma_2 = 100$ is our fixed budget, and $\gamma_1, \gamma_2 \in [0, 100]$.

Notice by observation of Fig. 5.2 that there is only a solution to the minimization problem when $\gamma_1 < \gamma_2$ such that $P_b^{opt} \in [0, 1)$. This may be proven mathematically as well, by using the fact that the cost function must have an inflection point. However, in order to complete the proof, we need to use the conjecture that $f_{CA}(P_b) = P(C, A)$ and $f_{C'A}(P_b) = P(C', A)$ are convex and monotonically decreasing. This is evident from Fig. 5.3, but much harder to prove formally.

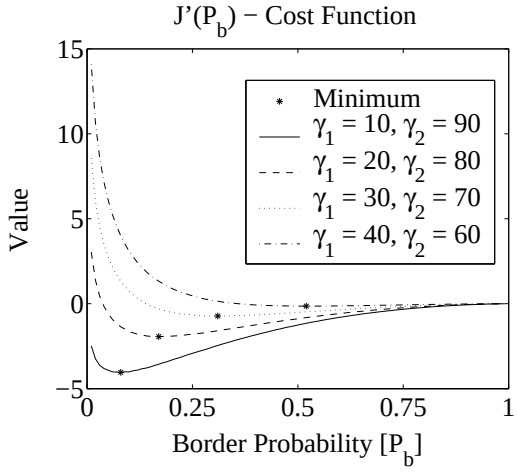

Figure 5.2: Solutions to $\arg\min_{P_b} J'(P_b)$ for differing γ_1, γ_2

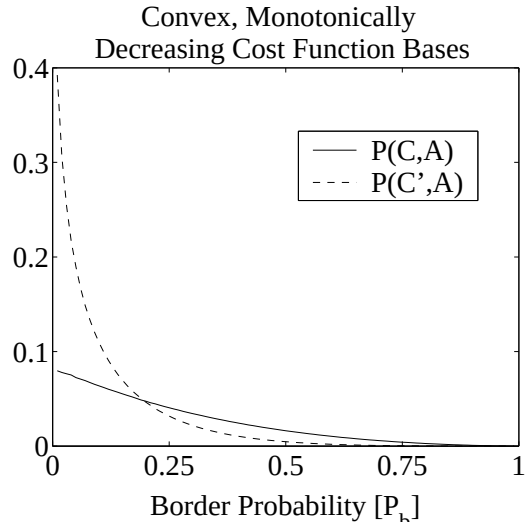

Figure 5.3: $f_{CA}(P_b) = P(C, A)$ and $f_{C'A}(P_b) = P(C', A)$

Because $\gamma_1 < \gamma_2$, this implies that there is a restriction on the cost of alarms in general to be less than the cost of **correct** alarms. Along with this restriction, γ_1 and γ_2 must still be based completely on heuristic selection. We can even modify the cost function, so that the cost weights are not "combined," as in the use of $\gamma'_{FA} = \gamma_{FA}P(A)$ and $\gamma'_{MD} = \gamma_{MD}P(A)$, but explicitly defined and penalizing their respective terms, as follows:

$$J(P_b) = \gamma_A f_A(P_b) + \gamma_{FA} f_{FA}(P_b) + \gamma_C P(C) + \gamma_{MD} f_{MD}(P_b) \quad (5.6)$$

$$\Rightarrow \quad J'(P_b) = \gamma_A f_A(P_b) + \gamma_{FA} f_{FA}(P_b) + \gamma_{MD} f_{MD}(P_b) \quad (5.7)$$

So we must now solve:

$$P_b^{opt} = \arg\min_{P_b} J(P_b) \quad (5.8)$$

$$= \arg\min_{P_b} J'(P_b) \quad (5.9)$$

$$= \arg\min_{P_b} \gamma_A f_A(P_b) + \gamma_{FA} f_{FA}(P_b) + \gamma_{MD} f_{MD}(P_b) \quad (5.10)$$

and choose γ_A, γ_{FA}, and γ_{MD} heuristically. However, now there are three free parameters for heuristic selection, and the cost function will become a more complicated function of P_b. In fact, there may be more complicated restrictions on choices of γ_A, γ_{FA}, and γ_{MD}, to allow for a solution to the minimization problem such that $P_b^{opt} \in [0, 1)$. This is due to the fact that more than one inflection point may exist, due to the additional free parameter.

Rather than using these types of cost functions that require the heuristic, restricted selection of weights, let us now introduce an alternative cost function. This cost function is more appropriate for our particular application target, because it is based upon the expected number of complaints over a time period instead of probabilities. Furthermore, it requires no heuristic selection of weights, and focuses on minimizing false alarms, which is more important for our application.

5.2.2 A New Optimal Alarm Design Method

One way to design the alarm system is based upon a metric that is very similar to probability of correct alarm, $P(C|A)$. However, instead of maximizing the probability of correct alarm, we quantify a new metric, the expected number of complaints conditioned on alarm, $E[C|A]$. Recall Examples 3 and 4 from Chap. 3. The probability of missed detection could not be lowered without

severely compromising the probability of false alarms. As such, the cost function we introduce in this chapter will allow us to focus on minimizing false alarms, since that has a higher priority for our application.

Minimizing false alarms achieves the same objective as maximizing correct alarms. Since we're dealing with **multiple** complaints, the objective can be reformulated even further to maximizing the expected number of complaints conditioned on an alarm, $E[C|A]$. Unlike the probability of false alarms, this can be expressed as a convex function of the border probability, and a unique maximum can be found. Furthermore, the computational burden does *not* scale exponentially in the number of time steps, hence there are no large contributions to excessive design time.

Its value as an alternative cost function in practice will be studied in earnest in Chap. 7. However, we will show that using this new cost function, $E[C|A]$, has favorable computational properties. Finding this metric as a function of P_b has a computational burden that scales linearly in the number of time steps under consideration. Therefore, unlike its counterpart cost functions based upon "basis functions," introduced in the previous section, the design time does not suffer from the curse of dimensionality. Although this is the case, we will find the integrals that need to be computed are quite complex. Specifically, the integrand of the numerator of the formula for $E[C|A]$ shown in Eqn. 5.11 is the product of two terms: one which is the standard multivariate normal probability distribution, and the other which quantifies the conditional expectation of the number of complaints.

$$E[C|A] = \frac{\int_{\Omega_A} \mathcal{N}(\hat{\mathbf{x}}; \mu_{\hat{\mathbf{x}}}, \Sigma_{\hat{\mathbf{x}}}) E[C|x_0, \ldots, x_k] d\hat{\mathbf{x}}}{P(A)} \tag{5.11}$$

Traditionally, we use a numerically robust algorithm developed by Genz [27], in order to compute multivariate normal probabilities. However, the algorithm can be modified slightly to accommodate the computation of expected values. In doing so, we are still faced with expressing and computing $E[C|x_0, \ldots, x_k]$ cleverly. Genz's algorithm is based on a Monte-Carlo sampling technique, so that the $E[C|x_0, \ldots, x_k]$ portion of the integrand entrenches itself in the algorithm on

the same level as the sampling. Therefore, with each new sample, the algorithm must also compute and use this value. If this portion of the integrand is not in closed form, then this can be very computationally expensive.

As we have shown in the previous chapter, the formula for $E[C|x_0, \ldots, x_k]$ in general does not have a closed form. Only when we are dealing with a zero-level, zero-mean stationary random process does this formula revert to closed form. Therefore, we must compute $E[C|x_0, \ldots, x_k]$ via Monte Carlo integration, at each sampling step of the original Monte Carlo integration that we were performing in the first place ! This has obvious disadvantages, both numerically and in terms of the computational burden. As such, we can cleverly apply an approximation introduced in the previous chapter in order to achieve a semi-parametric expression to reduce the computational burden, and avoid having to use "recursive" Monte-Carlo integrations.

Single Interval $2D$ Approximation

For the two-dimensional approximation with a single interval, the formula for the cost function shown in Eqns. 5.11 and 5.12 was derived using the rule of average conditional expectation [57]. The state-space model introduced in Chap. 3 is used to illustrate this concept, and all remaining approximations in this section.

$$
\begin{aligned}
E[C|A] &= \frac{\int_{\Omega_A} \mathcal{N}(\hat{\mathbf{x}}; \mu_{\hat{\mathbf{x}}}, \Sigma_{\hat{\mathbf{x}}}) E[C|x_0, \ldots, x_k] d\hat{\mathbf{x}}}{P(A)} \\
&= \frac{\int_{\Omega_A} \mathcal{N}(\hat{\mathbf{x}}; \mu_{\hat{\mathbf{x}}}, \Sigma_{\hat{\mathbf{x}}}) E[C|x_0, \ldots, x_k] d\hat{\mathbf{x}}}{\int_{\Omega_A} \mathcal{N}(\hat{\mathbf{x}}; \mu_{\hat{\mathbf{x}}}, \Sigma_{\hat{\mathbf{x}}}) d\hat{\mathbf{x}}} \\
\text{where } A &\triangleq \{\hat{\mathbf{x}} \in \Omega_A \subset \mathbb{R}^2\} \\
&= \{\hat{x}_{k|k} < \mathcal{X}_0, \hat{x}_{k+m|k} > \mathcal{Y}_m\} \\
\mathcal{X}_0 &= L - \sqrt{V_{k|k}} \Phi^{-1}(P_b) \\
\mathcal{Y}_m &= L + \sqrt{V_{k+m|k}} \Phi^{-1}(P_b) \\
\hat{\mathbf{x}} &= \begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+m|k} \end{bmatrix} \in \mathbb{R}^2 \\
\mu_{\hat{\mathbf{x}}} &= \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}(\mathbf{A}^m \mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^m)(\mathbf{I}_n - \mathbf{A})^{-1} \mathbf{B} \mu_{T_B}) \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \\
\Sigma_{\hat{\mathbf{x}}} &= \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A}^m \end{bmatrix} (\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R) \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A}^m \end{bmatrix}^T \\
\Rightarrow E[C|A] &= \frac{\int_{-\infty}^{\mathcal{X}_0} \int_{\mathcal{Y}_m}^{\infty} \mathcal{N}(\hat{\mathbf{x}}; \mu_{\hat{\mathbf{x}}}, \Sigma_{\hat{\mathbf{x}}}) E[C|x_0, \ldots, x_k] d\hat{\mathbf{x}}}{\int_{-\infty}^{\mathcal{X}_0} \int_{\mathcal{Y}_m}^{\infty} \mathcal{N}(\hat{\mathbf{x}}; \mu_{\hat{\mathbf{x}}}, \Sigma_{\hat{\mathbf{x}}}) d\hat{\mathbf{x}}} \qquad (5.12)
\end{aligned}
$$

We have not yet addressed the issue of the portion of the integrand represented by $E[C|x_0, \ldots, x_k]$. Unfortunately, if we use it in the form introduced in the previous chapter with Eqn. 5.12, we obtain the expression below.

$$
\begin{aligned}
E[C|A] &= \frac{\int_{-\infty}^{\mathcal{X}_0} \int_{\mathcal{Y}_m}^{\infty} \mathcal{N}(\hat{\mathbf{x}}; \mu_{\hat{\mathbf{x}}}, \Sigma_{\hat{\mathbf{x}}}) \left[\sum_{i=0}^{m-1} \int_{-\infty}^{L} \int_{L}^{\infty} \mathcal{N}(\mathbf{x}_i; \mu_{\mathbf{x}_i}, \Sigma_{\mathbf{x}_i}) d\mathbf{x}_i \right] d\hat{\mathbf{x}}}{\int_{-\infty}^{\mathcal{X}_0} \int_{\mathcal{Y}_m}^{\infty} \mathcal{N}(\hat{\mathbf{x}}; \mu_{\hat{\mathbf{x}}}, \Sigma_{\hat{\mathbf{x}}}) d\hat{\mathbf{x}}} \\
\text{Recall } \mu_{\mathbf{x}_i} &= \begin{bmatrix} \hat{x}_{k+i|k} \\ \hat{x}_{k+i+1|k} \end{bmatrix} \text{ and } \hat{\mathbf{x}} = \begin{bmatrix} \hat{x}_{k|k} \\ \hat{x}_{k+m|k} \end{bmatrix} \in \mathbb{R}^2 \\
i \in [0, m-1] &\Rightarrow \mu_{\mathbf{x}_i} \not\subseteq \hat{\mathbf{x}} \qquad (5.13)
\end{aligned}
$$

Therefore, this integral cannot be computed, due to the condition in Eqn. 5.13. However,

the case above is for the two-dimensional approximation with a single interval. If we choose any of the remaining optimal approximation methods, we may be able to satisfy a similar condition.

Multiple Sub-Interval Approximation

When we choose the multiple sub-interval approximation method, we can compute the expected number of complaints conditioned on an alarm over the entire prediction window according to the formula in Eqn. 5.14. However, this is only when the number of sub-intervals is equivalent to the number of steps in the prediction window, $N_s = m$. Also notice that the inclusion/exclusion rule is *not* used here, which is the basis of reducing the computational burden over the "basis function" method. What warrants not using the inclusion/exclusion rule is that Eqn. 5.14 is not a probability computation, and furthermore $E[C_i|A_i]$ represents the "expected number of complaints in interval i conditioned on an alarm in interval i." Because the formula evaluates to a number of complaints rather than a probability, we simply add up the expected number of complaints for all intervals. In truth, this can be considered an extension to the approximation used with the multiple sub-interval method, because we can't express the real alarm region as the union of sub-interval regions as in Eqn. 3.43. So we perform an "approximation" of sorts by integrating over the alarm region as defined in a piecewise manner, by using Eqn. 5.14.

$$
\begin{aligned}
E[C|A] &= \sum_{i=1}^{N_s} E[C_i|A_i] \qquad (5.14)\\
\text{where } E[C_i|A_i] &= \frac{\int_{\Omega_{A_i}} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) E[C_i|x_0, \ldots, x_k] d\hat{\mathbf{x}}_i}{P(A_i)} \\
&= \frac{\int_{\Omega_{A_i}} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) E[C_i|x_0, \ldots, x_k] d\hat{\mathbf{x}}_i}{\int_{\Omega_{A_i}} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) d\hat{\mathbf{x}}_i}
\end{aligned}
$$

$$
\begin{aligned}
\text{where } A_i &\triangleq \{\hat{\mathbf{x}}_i \in \Omega_{A_i} \subset \mathbb{R}^2\} \\
&= \{\hat{x}_{k+s_i|k} < \mathcal{X}_{s_i}, \hat{x}_{k+s_{i+1}|k} > \mathcal{Y}_{s_{i+1}}\} \\
\mathcal{X}_{s_i} &= L - \sqrt{V_{k+s_i|k}}\Phi^{-1}(P_b) \\
\mathcal{Y}_{s_{i+1}} &= L + \sqrt{V_{k+s_{i+1}|k}}\Phi^{-1}(P_b) \\
0 \leq s_i, s_{i+1} \leq m &\Leftrightarrow \text{indices corresponding to endpoints of subinterval } i \\
\hat{\mathbf{x}}_i &= \begin{bmatrix} \hat{x}_{k+s_i|k} \\ \hat{x}_{k+s_{i+1}|k} \end{bmatrix} \in \mathbb{R}^2 \\
\mu_{\hat{\mathbf{x}}_i} &= \begin{bmatrix} \mathbf{C}(\mathbf{A}^{s_i}\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^{s_i})(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \\ \mathbf{C}(\mathbf{A}^{s_{i+1}}\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^{s_{i+1}})(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \end{bmatrix} \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \\
\Sigma_{\hat{\mathbf{x}}_i} &= \begin{bmatrix} \mathbf{C}\mathbf{A}^{s_i} \\ \mathbf{C}\mathbf{A}^{s_{i+1}} \end{bmatrix} (\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R}) \begin{bmatrix} \mathbf{C}\mathbf{A}^{s_i} \\ \mathbf{C}\mathbf{A}^{s_{i+1}} \end{bmatrix}^T \\
\text{and } E[C_i|x_0, \ldots, x_k] &= P(x_{k+s_i} < L, x_{k+s_{i+1}} > L|x_0, \ldots, x_k) \qquad (5.15) \\
&= \int_{-\infty}^{L}\int_{L}^{\infty} \mathcal{N}(\mathbf{x}_i; \mu_{\mathbf{x}_i}, \Sigma_{\mathbf{x}_i}) d\mathbf{x}_i
\end{aligned}
$$

Note that Eqn. 5.15 only holds for the case when $\Delta_{s_i} \triangleq s_{i+1} - s_i = 1$, or in other words, when $N_s = m$. This is when the number of sub-intervals in question is the maximum possible, and the endpoints of all sub-intervals correspond to a pair of discrete time slices that are consecutive and adjacent. We see that this is a requirement, because in general, $E[C_i|x_0, \ldots, x_k] = P(x_{k+s_i} < L, x_{k+s_{i+1}} > L|x_0, \ldots, x_k)$ is not a valid formula for the conditional expectation for the number of complaints. **The expected number of complaints is equal to the probability only when we are considering the span of a single time interval.** This is true if we consider the event of interest as a Bernoulli random variable. With this assumption in place, we must redefine $\mathbf{x}_i; \mu_{\mathbf{x}_i}, \Sigma_{\mathbf{x}_i}$, as follows:

$$
\begin{aligned}
\mathbf{x}_i &= \begin{bmatrix} x_{k+s_i|k} \\ x_{k+s_{i+1}|k} \end{bmatrix} = \begin{bmatrix} x_{k+i-1}|x_0,\ldots,x_k \\ x_{k+i}|x_0,\ldots,x_k \end{bmatrix} \in \mathbb{R}^2 \\
\mu_{\mathbf{x}_i} &= \begin{bmatrix} \hat{x}_{k+s_i|k} \\ \hat{x}_{k+s_{i+1}|k} \end{bmatrix} = \begin{bmatrix} \hat{x}_{k+i-1|k} \\ \hat{x}_{k+i|k} \end{bmatrix} \\
\Sigma_{\mathbf{x}_i} &= \begin{bmatrix} Var(x_{k+s_i|k}) & cov(x_{k+s_i|k}, x_{k+s_{i+1}|k}) \\ cov(x_{k+s_{i+1}|k}, x_{k+s_i|k}) & Var(x_{k+s_{i+1}|k}) \end{bmatrix} \\
&= \begin{bmatrix} Var(x_{k+i-1|k}) & cov(x_{k+i-1|k}, x_{k+i|k}) \\ cov(x_{k+i|k}, x_{k+i-1|k}) & Var(x_{k+i|k}) \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
Var(x_{k+i-1|k}) &= \mathbf{C}[\mathbf{A}^{i-1}(\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R)(\mathbf{A}^T)^{i-1} + \mathbf{P}_{ss}^L]\mathbf{C}^T + \mathbf{R} \\
cov(x_{k+i-1|k}, x_{k+i|k}) &= \mathbf{C}[\mathbf{A}^{i-1}\hat{\mathbf{P}}_{ss}^R(\mathbf{A}^T)^i + \mathbf{L}_{ss} - \mathbf{A}^{i-1}\mathbf{L}_{ss}(\mathbf{A}^T)^{i-1}]\mathbf{C}^T \\
cov(x_{k+i|k}, x_{k+i-1|k}) &= \mathbf{C}[\mathbf{A}^{i}\hat{\mathbf{P}}_{ss}^R(\mathbf{A}^T)^{i-1} + \mathbf{L}_{ss} - \mathbf{A}^{i-1}\mathbf{L}_{ss}(\mathbf{A}^T)^{i-1}]\mathbf{C}^T \\
Var(x_{k+i|k}) &= \mathbf{C}[\mathbf{A}^{i}(\mathbf{P}_{ss}^L - \hat{\mathbf{P}}_{ss}^R)(\mathbf{A}^T)^{i} + \mathbf{P}_{ss}^L]\mathbf{C}^T + \mathbf{R} \\
\text{where } \mathbf{L}_{ss} &= \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T(\mathbf{A}^T)
\end{aligned}
$$

Therefore, to summarize the equations of interest for the new cost function metric when considering the sub-interval approximation ($N_s = m$), we have the following:

$$
\begin{aligned}
\Rightarrow\ E[C_i|A_i] &= \frac{\int_{-\infty}^{\mathcal{X}_{s_i}} \int_{\mathcal{Y}_{s_{i+1}}}^{\infty} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) \left[\int_{-\infty}^{L} \int_{L}^{\infty} \mathcal{N}(\mathbf{x}_i; \mu_{\mathbf{x}_i}, \Sigma_{\mathbf{x}_i}) d\mathbf{x}_i \right] d\hat{\mathbf{x}}_i}{\int_{-\infty}^{\mathcal{X}_{s_i}} \int_{\mathcal{Y}_{s_{i+1}}}^{\infty} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) d\hat{\mathbf{x}}_i} \\
\textbf{but now } \mu_{\mathbf{x}_i} &\equiv \hat{\mathbf{x}}_i \text{ !} \\
\Rightarrow\ E[C_i|A_i] &= \frac{\int_{-\infty}^{\mathcal{X}_{s_i}} \int_{\mathcal{Y}_{s_{i+1}}}^{\infty} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) \overbrace{\left[\int_{-\infty}^{L} \int_{L}^{\infty} \mathcal{N}(\mathbf{x}_i; \hat{\mathbf{x}}_i, \Sigma_{\mathbf{x}_i}) d\mathbf{x}_i \right]}^{g(\hat{\mathbf{x}}_i) \triangleq E[C_i|x_0,\ldots,x_k]} d\hat{\mathbf{x}}_i}{\int_{-\infty}^{\mathcal{X}_{s_i}} \int_{\mathcal{Y}_{s_{i+1}}}^{\infty} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) d\hat{\mathbf{x}}_i} \\
&= \frac{\int_{-\infty}^{\mathcal{X}_{s_i}} \int_{\mathcal{Y}_{s_{i+1}}}^{\infty} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) g(\hat{\mathbf{x}}_i) d\hat{\mathbf{x}}_i}{\int_{-\infty}^{\mathcal{X}_{s_i}} \int_{\mathcal{Y}_{s_{i+1}}}^{\infty} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) d\hat{\mathbf{x}}_i} \\
\text{Finally, } E[C|A] &= \sum_{i=1}^{N_s} E[C_i|A_i] \\
&= \sum_{i=1}^{N_s} \frac{\int_{-\infty}^{\mathcal{X}_{s_i}} \int_{\mathcal{Y}_{s_{i+1}}}^{\infty} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) g(\hat{\mathbf{x}}_i) d\hat{\mathbf{x}}_i}{\int_{-\infty}^{\mathcal{X}_{s_i}} \int_{\mathcal{Y}_{s_{i+1}}}^{\infty} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) d\hat{\mathbf{x}}_i}
\end{aligned}
\tag{5.16}
$$

Eqn. 5.16 is the formula for the new cost function parameterized by P_b. Notice how it scales linearly in N_s. Unlike the probabilistic computation, there is no need for the inclusion/exclusion rule, and therefore we do not suffer from the curse of dimensionality.

Multi-Dimensional Approximation

By using the multi-dimensional approximation method, we can also compute the expected number of complaints over the entire prediction window according to the formula in Eqn. 5.17. The formula is computed in consideration of the null hypothesis that tests for operating complaints only, as defined in Chap. 3, i.e. for Eqn. 3.53. The null hypothesis that tests for both arrival and operating complaints can be computed similarly, but the formulae are omitted. Again, as before with the multiple sub-interval approximation, computation of $E[C|A]$ in Eqn. 5.17 does not invoke the inlucsion/exclusion rule for the same reasons as stated prior.

Furthermore, the true approximate alarm region as shown in Eqn. 3.53 is implemented via an extended approximation as before with the multiple sub-interval method. The manner of implementing this extended approximation is shown in Eqn. 5.17 with piecewise adjacent alarm

regions, A_{i-1}, A_i. The adjacency is needed because multi-dimensional approximation alarm regions are associated with a single time slice, A_i. The integrand of the formula for the numerator of $E[C_i|A_{i-1}, A_i]$, which is $E[C_i|x_0, \ldots, x_k]$, will be shown to be a function of both $\hat{x}_{k+i-1|k}$ and $\hat{x}_{k+i|k}$. As a result, in order to integrate over a function of these two variables/time slices, adjacent alarm regions defining the corresponding regions of integration are required. This same logic does not apply for the multiple sub-interval approximation method because a single alarm region is defined with two adjacent time slices.

Recall that the example shown below is for in consideration of the null hypothesis that tests for operating complaints only, as defined in Chap. 3, i.e. for Eqn. 3.53: $A_0 \bigcap [\bigcup_{i=1}^{m} A_i]$. Consequently, the region of integration should *really* involve three time slices, A_0, A_{i-1}, A_i, however, this is not a particularly important issue due to the nature of the existing alarm region approximation. Furthermore, this same issue is not important when considering the null hypothesis that tests for both arrival and operating complaints.

$$
\begin{aligned}
E[C|A] &= \sum_{i=1}^{m} E[C_i|A_{i-1}, A_i] \qquad (5.17)\\
\text{where } E[C_i|A_{i-1}, A_i] &= \frac{\int_{\Omega_{A_{i-1},A_i}} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) E[C_i|x_0, \ldots, x_k] d\hat{\mathbf{x}}_i}{P(A_i)}\\
&= \frac{\int_{\Omega_{A_{i-1},A_i}} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) E[C_i|x_0, \ldots, x_k] d\hat{\mathbf{x}}_i}{\int_{\Omega_{A_i}} \mathcal{N}(\hat{\mathbf{x}}_i; \mu_{\hat{\mathbf{x}}_i}, \Sigma_{\hat{\mathbf{x}}_i}) d\hat{\mathbf{x}}_i}\\
A_0 &\triangleq \hat{x}_{k|k} \leq \overbrace{L - \sqrt{V_{k|k}}\Phi^{-1}(P_b)}^{\mathcal{X}_0}\\
A_i &\triangleq \hat{x}_{k+i|k} \geq \underbrace{L + \sqrt{V_{k+i|k}}\Phi^{-1}(P_b)}_{\mathcal{Y}_i}\\
\forall i &\in 1, \ldots, m
\end{aligned}
$$

Let's now make some clarifications of the terms used above:

$$
\begin{aligned}
\hat{\mathbf{x}}_i &= \begin{bmatrix} \hat{x}_{k+i-1|k} \\ \hat{x}_{k+i|k} \end{bmatrix} \in \mathbb{R}^2 \\
\mu_{\hat{\mathbf{x}}_i} &= \begin{bmatrix} \mathbf{C}(\mathbf{A}^{i-1}\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^{i-1})(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \\ \mathbf{C}(\mathbf{A}^{i}\mu_{\mathbf{q}} + (\mathbf{I}_n - \mathbf{A}^{i})(\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B}\mu_{T_B}) \end{bmatrix} \begin{bmatrix} \mathbf{C}\mu_{\mathbf{q}} \\ \mathbf{C}\mu_{\mathbf{q}} \end{bmatrix} \\
\Sigma_{\hat{\mathbf{x}}_i} &= \begin{bmatrix} \mathbf{C}\mathbf{A}^{i-1} \\ \mathbf{C}\mathbf{A}^{i} \end{bmatrix} (\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R}) \begin{bmatrix} \mathbf{C}\mathbf{A}^{i-1} \\ \mathbf{C}\mathbf{A}^{i} \end{bmatrix}^T \\
\text{and } E[C_i|x_0, \ldots, x_k] &= P(x_{k+i-1} < L, x_{k+i} > L | x_0, \ldots, x_k) \\
&= \int_{-\infty}^{L} \int_{L}^{\infty} \mathcal{N}(\mathbf{x}_i; \mu_{\mathbf{x}_i}, \Sigma_{\mathbf{x}_i}) d\mathbf{x}_i \qquad (5.18) \\
\text{where } \mathbf{x}_i &= \begin{bmatrix} x_{k+i-1|k} \\ x_{k+i|k} \end{bmatrix} \triangleq \begin{bmatrix} x_{k+i-1}|x_0, \ldots, x_k \\ x_{k+i}|x_0, \ldots, x_k \end{bmatrix} \in \mathbb{R}^2 \\
\mu_{\mathbf{x}_i} &= \begin{bmatrix} \hat{x}_{k+i-1|k} \\ \hat{x}_{k+i|k} \end{bmatrix} \\
\Sigma_{\mathbf{x}_i} &= \begin{bmatrix} Var(x_{k+i-1|k}) & cov(x_{k+i-1|k}, x_{k+i|k}) \\ cov(x_{k+i|k}, x_{k+i-1|k}) & Var(x_{k+i|k}) \end{bmatrix}
\end{aligned}
$$

Further clarification of the terms used above:

$$
\begin{aligned}
Var(x_{k+i-1|k}) &= \mathbf{C}[\mathbf{A}^{i-1}(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})(\mathbf{A}^T)^{i-1} + \mathbf{P}_{ss}^{L}]\mathbf{C}^T + \mathbf{R} \\
cov(x_{k+i-1|k}, x_{k+i|k}) &= \mathbf{C}[\mathbf{A}^{i-1}\hat{\mathbf{P}}_{ss}^{R}(\mathbf{A}^T)^{s_{i+1}} + \mathbf{L}_{ss} - \mathbf{A}^{i-1}\mathbf{L}_{ss}(\mathbf{A}^T)^{i-1}]\mathbf{C}^T \\
cov(x_{k+i|k}, x_{k+i-1|k}) &= \mathbf{C}[\mathbf{A}^{i}\hat{\mathbf{P}}_{ss}^{R}(\mathbf{A}^T)^{i-1} + \mathbf{L}_{ss} - \mathbf{A}^{i-1}\mathbf{L}_{ss}(\mathbf{A}^T)^{i-1}]\mathbf{C}^T \\
Var(x_{k+i|k}) &= \mathbf{C}[\mathbf{A}^{i}(\mathbf{P}_{ss}^{L} - \hat{\mathbf{P}}_{ss}^{R})(\mathbf{A}^T)^{i} + \mathbf{P}_{ss}^{L}]\mathbf{C}^T + \mathbf{R} \\
\text{where } \mathbf{L}_{ss} &= \mathbf{A}\mathbf{L}_{ss}\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T(\mathbf{A}^T)
\end{aligned}
$$

Therefore, to summarize the equations of interest for the new cost function metric when considering the multi-dimensional approximation, we have the following:

$$
\begin{aligned}
E[C_1|A_0,A_1] &= \frac{\int_{-\infty}^{\mathcal{X}_0}\int_{\mathcal{Y}_1}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})\left[\int_{-\infty}^{L}\int_{L}^{\infty}\mathcal{N}(\mathbf{x}_i;\mu_{\mathbf{x}_i},\Sigma_{\mathbf{x}_i})d\mathbf{x}_i\right]d\hat{\mathbf{x}}_i}{\int_{-\infty}^{\mathcal{X}_{s_i}}\int_{\mathcal{Y}_{s_{i+1}}}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})d\hat{\mathbf{x}}_i} \\
\text{and } E[C_i|A_{i-1},A_i] &= \frac{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})\left[\int_{-\infty}^{L}\int_{L}^{\infty}\mathcal{N}(\mathbf{x}_i;\mu_{\mathbf{x}_i},\Sigma_{\mathbf{x}_i})d\mathbf{x}_i\right]d\hat{\mathbf{x}}_i}{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})d\hat{\mathbf{x}}_i} \\
\forall i &\in \{2,\ldots,m\} \\
\textbf{Again, } \mu_{\mathbf{x}_i} &\equiv \hat{\mathbf{x}}_i
\end{aligned}
$$

Now making some definitions to reduce the complexity of the equations:

$$
\begin{aligned}
\Rightarrow\ E[C_1|A_0,A_1] &= \frac{\int_{-\infty}^{\mathcal{X}_0}\int_{\mathcal{Y}_1}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})\overbrace{\left[\int_{-\infty}^{L}\int_{L}^{\infty}\mathcal{N}(\mathbf{x}_i;\hat{\mathbf{x}}_i,\Sigma_{\mathbf{x}_i})d\mathbf{x}_i\right]}^{g(\hat{\mathbf{x}}_i)\triangleq E[C_i|x_0,\ldots,x_k]}d\hat{\mathbf{x}}_i}{\int_{-\infty}^{\mathcal{X}_0}\int_{\mathcal{Y}_1}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})d\hat{\mathbf{x}}_i} \\
&= \frac{\int_{-\infty}^{\mathcal{X}_0}\int_{\mathcal{Y}_1}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})g(\hat{\mathbf{x}}_i)d\hat{\mathbf{x}}_i}{\int_{-\infty}^{\mathcal{X}_0}\int_{\mathcal{Y}_1}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})d\hat{\mathbf{x}}_i} \\
\Rightarrow\ E[C_i|A_{i-1},A_i] &= \frac{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})\overbrace{\left[\int_{-\infty}^{L}\int_{L}^{\infty}\mathcal{N}(\mathbf{x}_i;\hat{\mathbf{x}}_i,\Sigma_{\mathbf{x}_i})d\mathbf{x}_i\right]}^{g(\hat{\mathbf{x}}_i)\triangleq E[C_i|x_0,\ldots,x_k]}d\hat{\mathbf{x}}_i}{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})d\hat{\mathbf{x}}_i} \qquad (5.19) \\
&= \frac{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})g(\hat{\mathbf{x}}_i)d\hat{\mathbf{x}}_i}{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})d\hat{\mathbf{x}}_i} \\
\forall i &\in \{2,\ldots,m\} \\
\text{Finally, } E[C|A] &= \sum_{i=1}^{m}E[C_i|A_{i-1},A_i] \\
&= \frac{\int_{-\infty}^{\mathcal{X}_0}\int_{\mathcal{Y}_1}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})g(\hat{\mathbf{x}}_i)d\hat{\mathbf{x}}_i}{\int_{-\infty}^{\mathcal{X}_0}\int_{\mathcal{Y}_1}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})d\hat{\mathbf{x}}_i} \\
&\quad + \sum_{i=2}^{m}\frac{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})g(\hat{\mathbf{x}}_i)d\hat{\mathbf{x}}_i}{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})d\hat{\mathbf{x}}_i} \qquad (5.20)
\end{aligned}
$$

Again, this new cost function scales *linearly*, now with the number of time steps under consideration, m, since we're dealing with the multi-dimensional approximation. For Eqns. 5.16 (sub-interval) and 5.20 (multi-dimensional), the integrand of all numerators is the product of $\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})$ and $g(\hat{\mathbf{x}}_i)$. The latter cannot be expressed explicitly as a function of $\hat{\mathbf{x}}_i$. Recall:

$$g(\hat{\mathbf{x}}_i) \triangleq E[C_i|x_0,\ldots,x_k] = \int_{-\infty}^{L}\int_{L}^{\infty}\mathcal{N}(\mathbf{x}_i;\hat{\mathbf{x}}_i,\Sigma_{\mathbf{x}_i})d\mathbf{x}_i$$

It is possible to compute Eqns. 5.16 and 5.20 this way, using the "recursive" Monte Carlo integration technique mentioned before, but at a huge computational disadvantage. One might consider that in the derivation provided previously, (See Eqn. 5.19) we might take one extra step, and move the "inner integration" of the numerator back out. Therefore, a four-dimensional integration would act across the entire integrand so that we have a product of multivariate normal PDF's in the numerator as shown in Eqn. 5.21.

$$\begin{aligned} E[C_i|A_{i-1},A_i] &= \frac{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})\overbrace{\left[\int_{-\infty}^{L}\int_{L}^{\infty}\mathcal{N}(\mathbf{x}_i;\hat{\mathbf{x}}_i,\Sigma_{\mathbf{x}_i})d\mathbf{x}_i\right]}^{g(\hat{\mathbf{x}}_i)\triangleq E[C_i|x_0,\ldots,x_k]}d\hat{\mathbf{x}}_i}{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})d\hat{\mathbf{x}}_i} \\ &= \frac{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\int_{-\infty}^{L}\int_{L}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})\mathcal{N}(\mathbf{x}_i;\hat{\mathbf{x}}_i,\Sigma_{\mathbf{x}_i})d\mathbf{x}_i d\hat{\mathbf{x}}_i}{\int_{\mathcal{Y}_{i-1}}^{\infty}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\hat{\mathbf{x}}_i;\mu_{\hat{\mathbf{x}}_i},\Sigma_{\hat{\mathbf{x}}_i})d\hat{\mathbf{x}}_i} \\ \forall i &\in \{2,\ldots,m\} \end{aligned} \tag{5.21}$$

Unfortunately, barring any mathematical technicalities that might otherwise prevent us from performing this integrand consolidation, the integrand of the numerator is still quite complicated. It is true that the product of two multivariate normals is proportional to a multivariate normal. However, the important thing to notice is that the integration is being performed over $\mathbf{x}_i$ as well as $\hat{\mathbf{x}}_i$. In this case, recalling the definition of $\mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})$ when dealing with multivariate normals, the variable of integration, $\mathbf{x}$ appears *before* the semicolon, *not* afterwards. The *parameters* of the multivariate normal, the mean, $\mu_{\mathbf{x}}$, and the covariance matrix, $\Sigma_{\mathbf{x}}$, appear afterwards.

Because $\hat{\mathbf{x}}_i$ is the both a variable of integration and the *mean* of the second multivariate normal, the integration is not a standard one that can be computed easily with Genz's algorithm. This algorithm was intended to be used for multivariate normal probability computations. The best alternative would be to use some other Monte Carlo technique to solve the integral. However,

this would require much more effort than necessary since there is already a robust technique to compute multivariate normal probabilities, specifically over semi-infinite half-spaces. Therefore, unless we use an approximation, the only other way to compute the portion of the integrand given by $E[C_i|x_0,\ldots,x_k]$ is to use Monte Carlo integration recursively as described previously, using the expression of $E[C_i|x_0,\ldots,x_k]$ shown below.

$$g(\hat{\mathbf{x}}_i) \triangleq E[C_i|x_0,\ldots,x_k] = \int_{-\infty}^{L}\int_{L}^{\infty} \mathcal{N}(\mathbf{x}_i;\hat{\mathbf{x}}_i,\Sigma_{\mathbf{x}_i})d\mathbf{x}_i$$

But instead, we can use the FOH discretization approximation revealed in the previous chapter, for level-crossings of a non-stationary Gaussian process in lieu of the above:

$$\begin{aligned} g(\hat{\mathbf{x}}_i) &\triangleq E[C_i|x_0,\ldots,x_k] \approx \left[\Phi\left(\frac{T-\mu}{s}\right) - \Phi\left(-\frac{\mu}{s}\right)\right]\left[\frac{\phi(\eta(\mu))}{\eta(\mu)} + \Phi(\eta(\mu))\right] \\ \text{where } \mu &= \frac{T(L-\hat{x}_{k+i-1|k})}{\hat{x}_{k+i|k}-\hat{x}_{k+i-1|k}} \\ \text{and } s &= \frac{1}{r'(\mu)} = \left[\frac{\sigma'(\mu)}{\sigma(\mu)}\left(\frac{m'(\mu)}{\sigma'(\mu)} - \frac{m(\mu)}{\sigma(\mu)}\right)\right]^{-1} \end{aligned}$$

$$\begin{aligned} \text{Recall } \phi(z) &= \text{Standard Normal PDF} = \frac{1}{\sqrt{2\pi}}\exp^{-\frac{z^2}{2}} \\ \text{and } \Phi(z) &= \text{Standard Normal CDF} = \int_{-\infty}^{z}\phi(\lambda)d\lambda \end{aligned}$$

$$\begin{aligned} \text{Furthermore } \eta(t) &= \frac{m' - \gamma(t)\varrho(t)r(t)}{\gamma(t)\sqrt{1-\varrho^2(t)}} \\ m' &\approx \frac{\hat{x}_{k+i|k}-\hat{x}_{k+i-1|k}}{T} \\ \gamma^2(t) &\approx \frac{\partial \mathbf{b}^T}{\partial t}\Sigma_{\mathbf{x}_i}\frac{\partial \mathbf{b}}{\partial t} \\ \varrho(t) &\approx \frac{\sigma'(t)}{\gamma(t)} = \frac{\frac{\partial \mathbf{b}^T}{\partial t}\Sigma_{\mathbf{x}_i}\mathbf{b}}{\sqrt{\mathbf{b}^T\Sigma_{\mathbf{x}_i}\mathbf{b}}} \\ r(t) &\approx \frac{\mathbf{b}^T\mu_{\mathbf{x}_i} - L}{\sqrt{\mathbf{b}^T\Sigma_{\mathbf{x}_i}\mathbf{b}}} \\ \text{and } \mathbf{b}^T &\triangleq \begin{bmatrix} 1-\frac{t}{T} & \frac{t}{T} \end{bmatrix} \end{aligned}$$

Because of the fact that $\{\hat{x}_{k+i-1|k}, \hat{x}_{k+i|k}\} \subset \hat{\mathbf{x}}_i$, the expression $g(\hat{\mathbf{x}}_i)$ can now be used as part of the integrand of Eqns. 5.16 and 5.20. These expressions are now much less computationally intensive, because $g(\hat{\mathbf{x}}_i)$ can now be expressed semi-parametrically as a function of $\hat{\mathbf{x}}_i$. In the next few sections, we will review and discuss the resulting new cost function, $E[C|A]$, that can be expressed as a function of P_b, for all cases of interest.

5.3 Examples

In order to demonstrate the results of the theory introduced in the previous section, we will use Examples 3 and 4 from Chap. 3, since they are of most practical interest to us. Therefore, we shall first consider the hypothesis of operating complaints only, Example 3, using the complaint model and relevant parameters. Recall that we use a 12-step ahead prediction window, or $m = 12$. This corresponds to a 4-hour period in which operating complaints can occur either during the morning or afternoon period of the day.

Figs. 5.4 and 5.5 show $E[C|A]$ plotted as a function of border probability, P_b, with respective P_b^{opt} maxima shown with asterisks. The dotted line represents the FOH approximation, and the solid line is the curve that was found by using the "recursive" Monte Carlo method. 3600 Monte Carlo Sampling points were used for the FOH approximations, and 360 were used in the recursive Monte Carlo method. Even though 10 times more sampling points were used, the FOH approximation results in a huge numerical and computational advantage. This is the case for all similar results shown in this section. Note that Fig. 5.4 shows the results when approximating the alarm region using the sub-interval method, and Fig. 5.5 show the results when approximation the alarm region with the multi-dimensional method.

The results for the second hypothesis in Example 4 from Chap. 3 are shown in Figs. 5.6, 5.7 and 5.8. The same complaint model and relevant parameters, as specified for both arrival and operating complaints are used. We also still use the same $m = 12$ step ahead prediction window as in

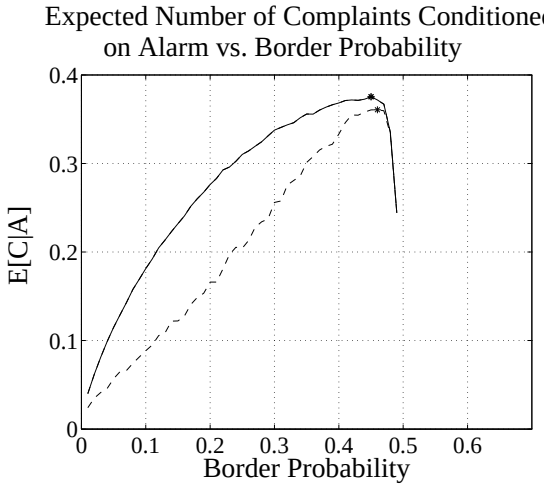

Figure 5.4: Sub-Interval Approximation, Operating Complaints Only (Example 3)

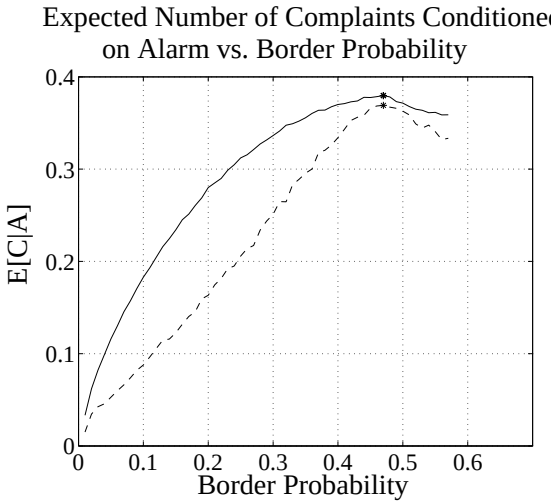

Figure 5.5: Multi-Dimensional Approximation, Operating Complaints Only (Example 3)

Example 3, and $d = 3$ as the number of steps prior to the start of the day. In this 1-hour period prior to the beginning of the day, we want to predict both arrival complaints and operating complaints in the ensuing 4-hour morning period. Fig. 5.6 shows the sub-interval approximation modelled with upcrossings, Fig. 5.7 shows the sub-interval approximation modelled with downcrossings, and Fig. 5.8 shows the multi-dimensional approximation.

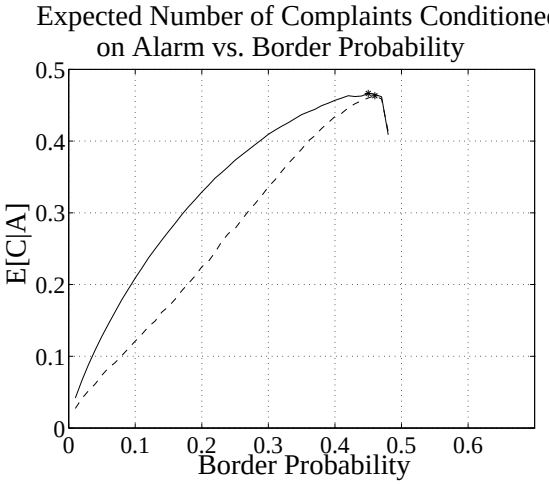

Figure 5.6: Sub-Interval Approximation, Both Arrival & Operating Complaints Only, Upcrossing Model (Example 4)

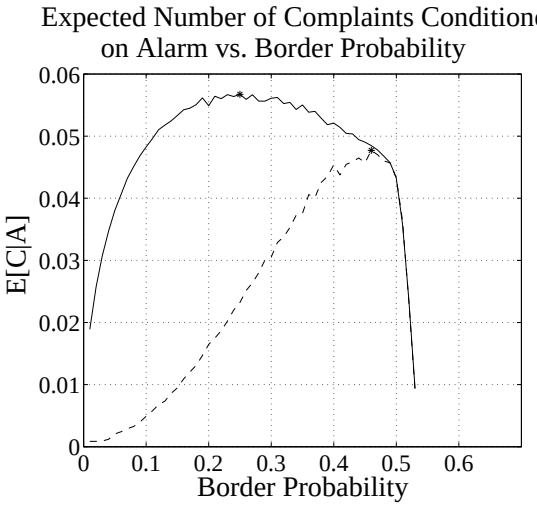

Figure 5.7: Sub-Interval Approximation, Both Arrival & Operating Complaints Only, Downcrossing Model (Example 4)

5.4 Discussion

Examining the results of the previous section, we see that the first-order hold (FOH) computation of $E[C|A]$ approximates the "recursive" Monte Carlo approach fairly well. Almost all of the graphs illustrate the characteristic that the maximum of the two cost functions shown

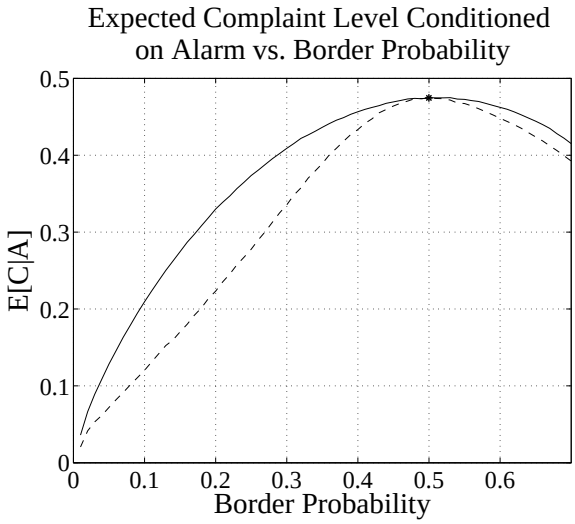

Figure 5.8: Multi-Dimensional Approximation, Both Arrival & Operating Complaints Only, (Example 4)

per figure are nearly identical, regardless of the computation method. However, on immediate inspection, it is clear that in all figures, the latter method shows more curvature. This can most likely be attributed to the fact that the FOH approximation is simply a linear interpolation.

Even when we are computing the $E[C|A]$ metric using complicated formulae such as Eqns. 5.16 and 5.20 as an obviously non-linear function of P_b, the "linear approximation" is still quite evident, specifically in Figs. 5.4, 5.5, 5.6, and 5.8. The only place that this behavior is not readily apparent is in the case where there is actually a distinction in the maximum between the two cost functions, in Fig. 5.7. The order of magnitude of the metric, $E[C|A]$ in Fig. 5.7, is also quite different the other graphs. The reason for this is unknown and requires further mathematical study. However, from a practical point of view, we can consider the sub-interval method *not* to be an approximation that should be used to obtain our optimal alarm operating point to minimize false alarms.

By looking at all of the results, we may surmise other some basic facts from a practical standpoint. For Example 3, operating complaints only, we see that using either the sub-interval (Fig. 5.4) or multi-dimensional (Fig. 5.5) approximation give almost the same answer in both

the maximum value and optimal P_b value. The optimal P_b values for the sub-interval and multi-dimensional methods are .45 and .47, respectively. For the sub-interval method, using the FOH approximation gives a slightly different answer of .46 vs. .45. The main qualitative difference between the sub-interval and multi-dimensional results for $E[C|A]$ is apparent after the maximum value in Figs. 5.4, 5.5. The reason for this can simply be attributed to the difference in the alarm regions between using the multi-dimensional method and the sub-interval method when $N_s = m = 12$.

Similarly, for Example 4, both arrival and operating complaints, we see that using either the sub-interval (upcrossing only, Fig. 5.6) or multi-dimensional (Fig. 5.8) approximation give close to the same answer in both the minimum value and P_b^{opt}. The optimal P_b values for the upcrossing sub-interval and multi-dimensional methods are .45 and .5, respectively. For the upcrossing sub-interval method, using the FOH approximation again gives a slightly different answer of .46 vs. .45. And again, the main qualitative difference between the sub-interval and multi-dimensional results for $E[C|A]$ is apparent after the maximum value in Figs. 5.6 and 5.8. However, due to the anomaly in Fig. 5.7 that illustrates the downcrossing method, the multi-dimensional method is always preferred to the sub-interval method, and at very little additional computational cost.

It is also interesting to note that the critical P_b^{opt} values found for the examples here correspond to a "critical" border probability when examining the probability-based results of Chap. 3. Specifically, the case of when we use upcrossing approximations for the alarm regions in Chap. 3 correspond to points at which there is a change in slope in Type I errors. This appears to be the same critical point in the $E[C|A]$-based cost functions in this chapter, where appropriate.

Aside from the huge numerical and computational advantage that there is to using the FOH approximation instead of the "recursive" Monte Carlo method of computing the new $E[C|A]$ cost function, there are other important practical considerations to review. Note that no heuristics are necessary for optimizing this new metric because there are no weights to select. Also, we want to maximize $E[C|A]$ as a function of P_b, as opposed to minimizing weighted cost functions as we did previously. Recall this minimization problem involved balancing the tradeoff between the fewest

possible alarms, and the greatest number of correct alarms, using heuristic weight selection. The new metric, $E[C|A]$, is very similar to probability of correct alarm, $P(C|A)$, which we'd like to maximize. The difference is that instead of maximizing $P(C|A)$ (probability of correct alarms), we are maximizing $E[C|A]$ (expected number of complaints given an alarm).

The design point found by optimizing the $E[C|A]$ cost function can easily be compared to the design point found by choosing a weighted combination of the probability based metrics. In fact, we can investigate the weights for the latter method that would result in the identical optimal design points for both methods. In other words, we want to find the equivalent costs of γ_1 and γ_2 that result in identical solutions to both optimization problems as in Eqns. 5.22, 5.23:

$$P_b^{opt} = \arg\min_{P_b} \gamma_1 f_A(P_b) - \gamma_2 f_{CA}(P_b) \tag{5.22}$$

$$= \arg\max_{P_b} E[C|A] \tag{5.23}$$

Suppose we choose an example with cost functions as shown in Fig. 5.2, based upon similar "basis functions" illustrated in Fig. 5.1. It is shown for several different cost weights of γ_1, and γ_2, where $\gamma_1 + \gamma_2 = 100$, and $\gamma_1, \gamma_2 \in [0, 100]$. Let us reformulate the basis functions somewhat so that they correspond to our particular cases of interest. Furthermore, let's assume that the total cost of $\gamma_1 + \gamma_2 = 100$ is a **monetary** cost of \$100.

First, let's consider the case when $P_b^{opt} = 0.45$. This corresponds to the optimal value of $E[C|A]$ for Example 3, whose alarm region is approximated by the sub-interval method when $N_s = m = 12$. The sub-interval method will provide the basis functions $f_A(P_b)$ and $f_{CA}(P_b)$ when $N_s = 4$ as a more computationally efficient and reasonable approximation to higher values of N_s due to the apparent convergence in Fig. 3.13. The resulting cost function weights that yield $P_b^{opt} = 0.45$ for the basis functions described are $\gamma_1 = \$7.5$ and $\gamma_2 = \$92.5$. Therefore, this implies that the cost of alarms in general are much less than the cost of **correct** alarms. In other words, we place much more of a premium on receiving spurious alarms than alarms in general. This is in qualitative

agreement with the same quantitative objective achieved by using the $E[C|A]$ design method.

Now, let's again consider the case when $P_b^{opt} = 0.45$, but this time corresponding to the optimal value of $E[C|A]$ for Example 4. This particular value of P_b^{opt} has an alarm region that is approximated by the "upcrossing" sub-interval method when $N_s = m = 12$. The sub-interval method will again provide the basis functions $f_A(P_b)$ and $f_{CA}(P_b)$ when $N_s = 4$ as a more computationally efficient approximation to higher values of N_s. The resulting cost function weights that yield $P_b^{opt} = 0.45$ for the basis functions described are now $\gamma_1 = \$36$ and $\gamma_2 = \$64$. Again this implies that we place more of a premium on receiving spurious alarms than alarms that are either correct or false, but to a lesser extent than when we are considering only operating complaints. Perhaps this has to do with the fact that arrival complaints have a different equivalent probabilistic interpretation than operating complaints.

At any rate, we know that these costs are only meant to serve as equivalences in the design of the alarm system based upon the $E[C|A]$ metric. If there were more concrete costs that could be assigned to alarms vs. false alarms, then we could use them in lieu of this metric. However, in the absence of such heuristically assigned costs, it may be reasonable to use this metric instead. The ramifications of doing so, as opposed to using other types of design criteria from a more concrete practical and simulation standpoint will be provided in Chap. 7.

5.5 Conclusion

The main conclusions to be derived from this short chapter can be summarized with the following bullets:

- The FOH approximation is much more computationally efficient than using the "recursive" Monte Carlo approach, and just as accurate in providing a solution to the optimal design problem of finding $P_b^{opt} = \arg\max_{P_b} E[C|A]$, for the upcrossing sub-interval and multi-dimensional approximations to the alarm regions. Practically, the multi-dimensional approximation is pre-

ferred to the sub-interval approximation design method.

- The new cost function, $E[C|A]$, is much more computationally efficient from a design standpoint than the cost functions that are based upon heuristically chosen weightings of basis functions for the probabilities $f_A(P_b) = P(A)$ and $f_{CA}(P_b) = P(C, A)$. The computation time of $E[C|A]$ scales linearly with the number of steps in the time interval rather than exponentially, as the basis functions do.

- The new cost function can easily be shown to provide an optimal value of P_b^{opt} that is equivalent to the value obtained by performing an optimization based upon the probabilistic "basis function" method. Doing so, we can find what the equivalent costs would be on alarms and false alarms, when using $E[C|A]$. As such, in the absence of such costs, using $E[C|A]$, or its FOH approximation is a much more computationally efficient alternative in the design of an optimal alarm system.

Chapter 6

Optimal Control

6.1 Introduction

The development of optimal control strategies for thermostat settings in commercial buildings has seen modest coverage in the literature [10, 12, 54]. Traditionally, optimal control algorithms are solved by minimizing an objective function similar to the one presented in Eqn. 2.4, where $J = \sum_{k=0}^{L} R_k P_k T_s$ is subject to some thermal comfort constraints. These constraints are typically derived from ASHRAE standards, such as ASHRAE Standard 55 shown in Table 2.2. The solution results in a trajectory of optimal zone temperature setpoints throughout the course of a 24-hour period. The methodology for solving these optimal control problems ranges from the use of dynamic programming to other types of constrained nonlinear optimization techniques. The primary goal of these algorithms is to minimize the dollar cost of energy expenditure (often including both time-of-use and demand energy charges), while providing an acceptable range of thermal comfort conditions.

It is of use to consider a variant of this problem, in which the cost of complaints is quantified, as in Eqn. 2.5. Here, $J = \sum_{k=0}^{L} R_k P_k T_s + C_k$, and C_k= the cost of complaints predicted by the complaint model introduced by Federspiel [19, 21]. In this case, it is not necessary to consider

thermal comfort as a constraint, because there is an additive term to characterize thermal comfort via the cost of complaints. It is assumed that each complaint is associated with the dispatch of a facility technician, therefore the response to them is unautomated. Details on computing the actual dollar cost of complaints by choice of labor rates, etc., are documented [21, 22].

The revised cost function is therefore time-varying and unconstrained, and not in quadratic form. The time-varying nature of the cost function can be derived from two sources. One source is from the energy term, which uses a time-varying price signal, R_k. This clearly has obvious implications for the burgeoning demand response research and enabling technology area in building controls. The second source derives from the fact that the complaint-only cost function is dependent on the time of day. Recall from Chap. 3 that arrival and operating complaints are distinctly categorized. This importance of this distinction is also made clear in work presented by Federspiel [19, 21, 22].

Due to the lack of quadratic form, it is not plausible to use a traditional LQ-optimal control approach to the problem as currently posed. Although the cost function is not quadratic, the plant under consideration is in fact linear, as it has been introduced and detailed in Chapter 3. However, it is not possible to use this model when considering the sum of both energy and complaint costs. This statistics-based model is based simply upon the fundamental spectral characteristics of actual building temperature time-series data, and therefore has no model of energy consumption by an actuator within a closed-loop feedback control system. To quantify energy costs, we need to use a more detailed model that is based upon first principles, and not purely with statistics, as has been introduced thus far. The model we will use is based on work that has presented previously by the current author [46, 47]. The state-space formulation that has been used for the statistics-based model does not need to be adjusted significantly to adopt the more complex first-principles based model.

The problem with using this more complex higher-dimensional state-space model is due to the *curse of dimensionality*. This has been verified by using the package of approximate dynamic

programming developed by Miranda and Fackler [49]. The curse of dimensionality is a common problem when employing dynamic programming to solve optimal control problems. The problem stems from the fact that the computational burden and storage requirements increase exponentially with the dimension of the system state. It has been found that the curse of dimensionality presents a problem even when penalizing only the complaint cost and not the energy cost at all, with a modest control horizon. This is true even considering the fact that we would still be relying on the 2^{nd} -order model that is derived purely from statistics. Even one additional dimension in the cost function (therefore a 3-state cost function) invokes the curse of dimensionality. Additional overhead from the chosen approximate dynamic programming method [49] no doubt incurs further computational burden. The extra dimension arises from the fact the complaint-only based cost function is time-varying for reasons stated previously.

A further use of the relevant statistics, but *without any model*, is to consider a purely static optimization for the computation of a nominal baseline thermostat setting. This optimization can be based upon the theory introduced in Chap. 4. Although the purpose of Sec. 4.3.2 was to introduce the approximation for use in a dynamic sense, it is just as possible to use these results for static optimization as well. In fact, as seen in Fig. 4.5, the optimal nominal baseline thermostat can just as well be determined by use of the continuous cost function without even having to perform a discretization in the static case. This is true because we are not seeking dynamic implementation based upon the complaint cost as a function of a dynamic parameter, but rather a static mean. Furthermore, it is also possible to perform a static optimization when considering energy costs as well, as demonstrated by Federspiel [19, 21, 22].

Therefore, we will modify the cost function to in order to address the optimal control question with a revised paradigm. Instead of developing a trajectory of optimal zone temperature setpoints throughout the course of a 24-hour period, we will construct an *optimal reference modification* policy. Under the previous paradigm, from a control theory point of view, the trajectory of optimal zone temperature setpoints acts as feedback control in an autonomously acting closed-loop

system. This is not specifically the method of implementation of previous studies as mentioned earlier [10, 12, 54], but rather a natural interpretation of this paradigm when implemented on-line in real-time. In a sense, the feedback control would close the loop, and inherently create a bang-bang nonlinear optimal controller, whose stability and performance characteristics would need much further study. A simple diagram of such a closed-loop system has been presented in Figure 1.1, and its autonomous counterpart is shown in Fig. 6.1.

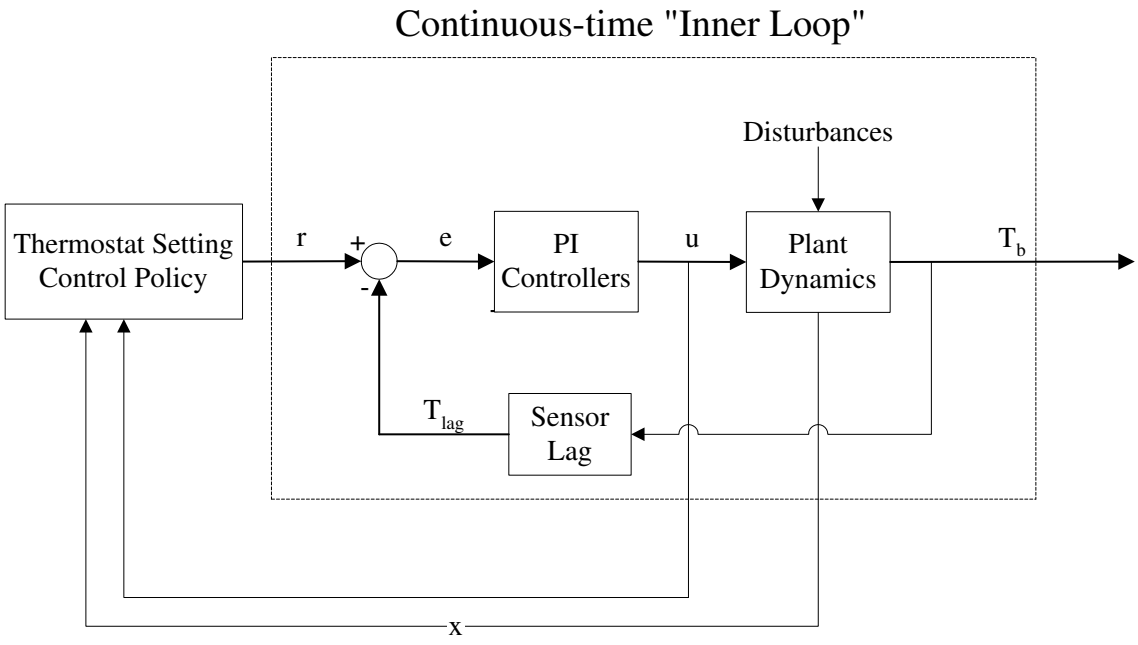

Figure 6.1: Block Diagram of Autonomous System

However, under the new paradigm, we consider response to individual complaint or alarm-based *events*. The cost function does not include the additive complaint cost term, but only the square cost of energy conditioned on the type of event being responded to. In the case of alarm events, we can think of this chapter as tying together two major themes of this thesis. One theme is all of the work on optimal alarm theory presented thus far, with the second theme of how to respond to them in an optimal fashion. We do this by combining optimal alarm theory with basic optimal control theory in this chapter. This is therefore the first presentation of a new "optimal alarm-based control," which

potentially can have major implications for any number of applications where automated decision and control based upon alarm signals is desired. In a sense, "alarm-based control" is an algorithm that will take an optimal preemptive action to prevent future complaint events in the most energy-efficient manner, based upon optimal prediction and alarm results. Therefore, the action is predictive rather than corrective, as in response to complaint events that have already occurred. However, we also consider corrective response to complaint events in this chapter.

The resulting control can be considered an *optimal open-loop control law*, rather than a closed-loop feedback control. This is due in part to the response to events on a case-by-case basis, but also due to the nature of the optimal control theory used to derive the algorithm, that cannot be interpreted as closed-loop optimal feedback control. The theory that follows in the subsequent section is based upon work by Hansson [32]. Hansson's work has also extended more generally to response to critical events using the closed-loop as well as open-loop control paradigm [30, 31, 32, 33]. However, with his closed-loop control scheme, "inner loop" controllers are designed to prevent future level-crossing events with feedback rather than acting at a supervisory level with reference modification. There are several advantages in using this new paradigm, listed as follows:

- The formulation can now be treated as an LQ optimal control algorithm, with appropriate approximations in place.
- The resulting control law is not autonomously acting and therefore there is no possibility for a bang-bang nonlinear control law which may cause wild fluctuations in temperature to maintain the control objective.
- The control uses an estimate of the state of the plant and system at the time of execution, when the event occurs, and therefore is more suitable in an open-loop context.
- Cleanly integrates the use of state information at the time of *any* type of critical event, whether it is an alarm *or* a complaint event. With closed-loop control, this distinction is not made.

6.2 Optimal Reference Modification

6.2.1 State-Space Representation of Plant

Recall from the introduction that the state-space formulation used for the statistics-based model does not need to be adjusted significantly to adopt the more complex first-principles based model. In fact, we may use the exact same definition of the building temperature process as before in Chap. 3 in continuous time:

$$\begin{aligned} \dot{\mathbf{q}}(t) &= \mathbf{A}\mathbf{q}(t) + \mathbf{B}u(t) + \mathbf{B}_w\mathbf{w}(t) \\ x(t) &= \mathbf{C}\mathbf{q}(t) + v(t) \end{aligned}$$

The main differences are now the state of the system, $\mathbf{q}(t)$, which represents the closed-loop state, and the input to the system, $u(t)$, which now should be replaced by $r(t) + u(t)$. $r(t)$ will now act as the scalar fixed control input, and $u(t)$ will act as the modification to that input. The closed-loop state is comprised of the building state ($\mathbf{q}_b(t) \in \mathbb{R}^n$), as defined by the details of the model from [46, 47], the controller state ($q_c(t) \in \mathbb{R}$) , and the sensor state ($q_s(t) \in \mathbb{R}$) such that

$$\mathbf{q}(t) \equiv \begin{bmatrix} \mathbf{q}_b(t) \\ q_c(t) \\ q_s(t) \end{bmatrix} \in \mathbb{R}^{n+2} \tag{6.1}$$

Furthermore, another difference is that in this model, we make no assumptions about the Gaussianity or stationarity of the input disturbance vector, $\mathbf{w}(t) \in \mathbb{R}^{m_d}$, where $m_d = 2$. The input disturbance is now a vector comprised of the outside temperature, and internal heat gains due to building occupants and equipment. Therefore, when each equation corresponding to the individual components of the closed-loop state equation as shown in Eqn. 6.1 are discretized we obtain the following:

$$\mathbf{q}_{b_{k+1}} = \mathbf{A}_{b_d}\mathbf{q}_{b_k} + \mathbf{B}_{b_d}h_k + \mathbf{B}_{w_{b_d}}\mathbf{w}_k \tag{6.2}$$

$$q_{c_{k+1}} = q_{c_k} + T_s e_k \tag{6.3}$$

$$q_{s_{k+1}} = A_{s_d}q_{s_k} + B_{s_d}x_k \tag{6.4}$$

Some important internal variables and outputs of the closed-loop system are as follows:

$$\begin{aligned} h_k &= K_i q_{c_k} + K_p e_k \\ x_{b_k} &= \mathbf{C}_{b_d}\mathbf{q}_k \\ x_k &= C_{s_d}q_{s_k} + v_k \\ e_k &= r_k + u_k - x_k \end{aligned}$$

In the above, e_k represents the control system error, x_{b_k} represents the actual building temperature, and x_k represents the temperature measured by the thermostat (including sensor measurement lag and noise). Notice a new output h_k: this represents the power consumed by the control actuator, which we are interested in quantifying for penalization in the cost function. This "inner loop" controller is a standard PI controller with control gains K_p and K_i. When equations 6.2-6.4 are concatenated as in Eqn. 6.1, the following state-space equation results:

$$\mathbf{q}_{k+1} = \mathbf{A}_d\mathbf{q}_k + \mathbf{B}_d(r_k + u_k) + \mathbf{B}_{w_d}\mathbf{w}_k + \mathbf{B}_{v_d}v_k \tag{6.5}$$

$$x_k = \mathbf{C}_d\mathbf{q}_k + v_k \tag{6.6}$$

$$p_k = \mathbf{M}_d\mathbf{q}_k \tag{6.7}$$

where

$$\mathbf{A}_d = \begin{bmatrix} \mathbf{A}_{b_d} & \mathbf{B}_{b_d}K_i & -\mathbf{B}_{b_d}K_pC_{s_d} \\ \mathbf{0}_{1xn} & 1 & -T_sC_{s_d} \\ B_{s_d}\mathbf{C}_{b_d} & 0 & A_{s_d} \end{bmatrix}, \mathbf{B}_d = \begin{bmatrix} \mathbf{B}_{b_d}K_p \\ T_s \\ 0 \end{bmatrix}$$

$$\mathbf{B}_{w_d} = \begin{bmatrix} \mathbf{B}_{w_{b_d}} \\ \mathbf{0}_{1xm_d} \\ \mathbf{0}_{1xm_d} \end{bmatrix}, \mathbf{B}_{v_d} = \begin{bmatrix} -\mathbf{B}_{b_d} K_p \\ -T_s \\ 0 \end{bmatrix}$$

and

$$\begin{aligned} \mathbf{C}_d &= \begin{bmatrix} \mathbf{0}_{1xm_d} & 0 & C_{s_d} \end{bmatrix} \\ \mathbf{M}_d &= \begin{bmatrix} \mathbf{0}_{1xn} & 1 & 0 \end{bmatrix} \\ v_k &\sim \mathcal{N}(0, R) \end{aligned}$$

Note that p_k represents the integral term in the PI controller, which is also equivalent to the state of the controller, q_{c_k}. This is necessary to compute the optimal control law when penalizing energy usage.

6.2.2 Propagation of Means of State Space Equations

When formulating the optimal control law, it is also of importance to obtain the means of the state space equations, and their conditional means. When conditioning on events, we will use the notation $[\cdot|C]$ for a complaint event, and $[\cdot|A]$ for an alarm event. When there is no need for distinction between the two, we will use the $[\cdot|\mathcal{E}]$ notation. The benefit of performing this will become more clear in the next few sections, but the basic method has been derived from Hansson [32]. Therefore, let's make a few definitions for clarity of presentation:

$$
\begin{aligned}
\bar{\mathbf{q}}_k &\triangleq E[\mathbf{q}_k] \\
\bar{x}_k &\triangleq E[x_k] \\
\bar{p}_k &\triangleq E[p_k] \\
r_k &= E[r_k] = E[r_k|\mathcal{E}] = \bar{r} \\
u_k &= E[u_k] = E[u_k|\mathcal{E}] \\
u_k^2 &= E[u_k^2] = E[u_k^2|\mathcal{E}] \\
\bar{w}_k &\triangleq E[e_k] = E[w_k|\mathcal{E}]
\end{aligned}
$$

Now taking the expectation of Eqn. 6.5 and applying the definitions above, we obtain the following:

$$\bar{\mathbf{q}}_{k+1} = \mathbf{A}_d\bar{\mathbf{q}}_k + \mathbf{B}_d(r_k + u_k) + \mathbf{B}_{w_d}\bar{\mathbf{w}}_k \tag{6.8}$$

$$\bar{x}_k = \mathbf{C}_d\bar{\mathbf{q}}_k \tag{6.9}$$

$$\bar{p}_k = \mathbf{M}_d\bar{\mathbf{q}}_k \tag{6.10}$$

For computation of the optimal control law, it will also be necessary to find the propagation of the state and output *error* vectors, defined as follows:

$$
\begin{aligned}
\tilde{q}_k &\triangleq q_k - \bar{q}_k \\
\tilde{x}_k &\triangleq x_k - \bar{x}_k \\
\tilde{p}_k &\triangleq p_k - \bar{p}_k \\
\tilde{w}_k &\triangleq w_k - \bar{w}_k
\end{aligned}
$$

Therefore,

$$\tilde{\mathbf{q}}_{k+1} = \mathbf{A}_d\tilde{\mathbf{q}}_k + \mathbf{B}_{w_d}\tilde{\mathbf{w}}_k \tag{6.11}$$

$$\tilde{x}_k = \mathbf{C}_d\tilde{\mathbf{q}}_k + v_k \tag{6.12}$$

$$\tilde{p}_k = \mathbf{M}_d\tilde{\mathbf{q}}_k \tag{6.13}$$

Finally, we must take the expectation of Eqn. 6.8, conditioned on an event, and make a few last definitions.

$$\begin{aligned}
\hat{\mathbf{q}}_{k|\mathcal{E}} &\triangleq E[\bar{\mathbf{q}}_k|\mathcal{E}] \\
\hat{x}_{k|\mathcal{E}} &\triangleq E[\bar{x}_k|\mathcal{E}] \Rightarrow \hat{x}^2_{k|\mathcal{E}} = E[\bar{x}^2_k|\mathcal{E}] \\
\hat{p}_{k|\mathcal{E}} &\triangleq E[\bar{p}_k|\mathcal{E}] \Rightarrow \hat{p}^2_{k|\mathcal{E}} = E[\bar{p}^2_k|\mathcal{E}] \\
\hat{\alpha}_{k|\mathcal{E}} &\triangleq E[\tilde{p}_k|\mathcal{E}] \\
\hat{\beta}_{k|\mathcal{E}} &\triangleq E[\tilde{p}^2_k|\mathcal{E}] \\
\hat{\xi}_{k|\mathcal{E}} &\triangleq E[\tilde{x}_k|\mathcal{E}] \\
\hat{\eta}_{k|\mathcal{E}} &\triangleq E[\tilde{x}^2_k|\mathcal{E}]
\end{aligned}$$

The $\Rightarrow$ shown above can be proven by the assumption of stationarity. The state equations can now be written as follows:

$$\hat{\mathbf{q}}_{k+1|\mathcal{E}} = f(\hat{\mathbf{q}}_{k|\mathcal{E}}, u_k, k) = \mathbf{A}_d\hat{\mathbf{q}}_{k|\mathcal{E}} + \mathbf{B}_d(r_k + u_k) + \mathbf{B}_{w_d}\bar{\mathbf{w}}_k \tag{6.14}$$

$$\hat{x}_{k|\mathcal{E}} = \mathbf{C}_d\hat{\mathbf{q}}_{k|\mathcal{E}} \tag{6.15}$$

$$\hat{p}_{k|\mathcal{E}} = \mathbf{M}_d\hat{\mathbf{q}}_{k|\mathcal{E}} \tag{6.16}$$

6.2.3 Revisions to Predictions and Optimal Alarm System

With the revisions in the state-space representation of the plant, all Kalman filtering/prediction and optimal alarm system design will be affected. Because of the change in the size of the state and

the input disturbances, certain changes need to be clarified. The main difference, as highlighted before, was in Eqn. 6.5. In order to obtain a true first-principles based optimal alarm system, we must use this same equation as the basis for all Kalman filtering and prediction that takes place. Recall that the use of the standard Kalman filter makes tacit assumptions about the Gaussian nature of input and measurement noise. Making a stationary Gaussian assumption about the measurement noise is certainly appropriate, since we've performed a study on actual measurement noise from real data in Sec. 3.3.2. However, it is more appropriate to choose measurement noise based upon matching the model's output to real data than based upon the measurements made in Sec. 3.3.2. It was found that we should use $R = 0.005$ in lieu of $R = 0.023759$, although the stationary Gaussian assumption about the measurement noise is still valid.

The input noise needs to be treated differently. The two disturbances that comprise this vector, $\mathbf{w}_k$, are inherently non-stationary, and non-Gaussian. Recall that they are the outside temperature (TMY data, T_{out}), and internal heat gains due to building occupants and equipment ($\dot{Q}_{int}$). In order to make any Gaussian assumptions about the input noise, we must first "remove the non-stationary mean" of the annual disturbance profile. In the case of the outside temperature, this mean is obtained by filtering interpolated TMY weather data with an exponentially-weighted moving average (EWMA) filter that has a very small forgetting factor, $\lambda = 0.001$. Although presented in a completely different context, the details of the EMWA filter have already been provided in Eqns. 2.12-2.13. The basic formula that is used to model the internal gains is presented in Eqn. 6.17. It is basically modelled as the linear combination of randomly phased and simulated building occupants and equipment within a zone, based on work by Lin and Federspiel [44]. The phasing is based on an on/off square wave with a randomized duty cycle, and the night/weekend periods are given residual background heat with a uniformly randomized component. This background heat is also present during the day.

$$\begin{aligned}
\mathbf{w}_k &= \begin{bmatrix} \dot{Q}_{int} \\ T_{out} \end{bmatrix} \\
\Rightarrow \dot{Q}_{int} &= B_0(1+0.1\mathbb{U}) + \sum_{i=1}^{N_c} \frac{C_0}{2}\left[1+s(\omega t_k+\phi_i)\right] + \sum_{i=1}^{N_p} \frac{B_0}{2}\left[1+s(\omega t_k+\phi_i)\right] \qquad (6.17)
\end{aligned}$$

where

$$\begin{aligned}
\omega &= \frac{2\pi T_s}{N} \\
N &= \text{Max on-time} = 8 \text{ hrs.} \\
\phi_i &= N_p\mathbb{U}(-1)^{\lfloor \mathbb{U} \rceil} \\
s(\delta) &= \text{Square Wave} = \begin{cases} 1 & \delta \in \mathbb{U} \times 100 \text{ duty cycle} \\ -1 & \delta \in (1-\mathbb{U}) \times 100 \text{ duty cycle} \end{cases} \\
N_c &= \text{Number of computers/equipment pieces in zone} \\
N_p &= \text{Number of building occupants in zone} \\
B_0 &= \text{Typical heat generated by a single building occupant} = 100 \text{ Watts} \\
C_0 &= \text{Typical heat generated by a single computer} = 140 \text{ Watts} \\
\mathbb{U} &= \text{random number from uniform distribution}
\end{aligned}$$

It is possible to compute the expected internal heat gain throughout the course of a year, which on nights and evenings is $1.05B_0$, and for all other periods is $1.05B_0 + \frac{C_0N_c+B_0N_p}{2}$. Now that it is clear how to compute the non-stationary mean, $\bar{\mathbf{w}}_k$, for both components of the input disturbance $\mathbf{w}_k$, it is now possible to compute its covariance matrix. It is defined as $\mathbf{W} = VC(\mathbf{w}_k)$, and computed by using Eqn. 6.18, which is an empirical maximum likelihood estimate of the covariance estimate based upon the Gaussian assumption, provided in Jordan [37].

$$\begin{aligned}
\mathbf{W} &= VC(\mathbf{w}_k) \\
&= \frac{\sum_{k=0}^{N}(\mathbf{w}_k - \bar{\mathbf{w}}_k)(\mathbf{w}_k - \bar{\mathbf{w}}_k)^T}{N} \qquad (6.18)
\end{aligned}$$

The estimate is based upon an empirical sample of N simulation steps, amounting to an annual profile of TMY weather and internal gain simulation and computation of non-stationary means, $\bar{\mathbf{w}}_k$. Therefore, to account for the mean and variance of the non-stationary and non-Gaussian input disturbance using the Gaussian paradigm, we can assume that $\mathbf{w}_k \sim \mathcal{N}(\bar{\mathbf{w}}_k, \mathbf{W})$. By subtracting off the non-stationary portion of the mean, $\bar{\mathbf{w}}_k$ to compute the covariance matrix, we are developing as accurate a Gaussian representation of the input noise as possible.

There is one final measure that needs to be taken to ensure that the Kalman filtering and prediction can be performed with the first-principles model to meet the Gaussian assumptions. It has to do with the second term of Eqn. 6.5, $\mathbf{B}_d(r_k + u_k)$. We know that r_k is a static reference setting, and u_k is its modification. Therefore, as far as these two terms' effect on Kalman filtering and Gaussianity, the first term is negligible. However for the second term, we can artificially treat the model's non-stationary night and weekend periods as stationary periods. Specifically, we can account for this by assuming that $u_k \sim \mathcal{N}(0, \sigma^2_{T_{out}})$. Because the building temperature is allowed to float on nights and weekends, the general trend is for it to follow the outside temperature during these periods. Therefore, we assume that the variance of the outside temperature is equivalent to the variance of the reference modification during these periods. The zero mean assumption on u_k is used because this is only a reference modification, not the actual static reference, which has no variance.

Given this, when simulating the *real* model, or running the real system as it propagates forward according to Eqn. 6.5, the Gaussian assumptions about u_k and $\mathbf{w}_k$ are unnecessary. The simulation will simply march forward according to the available TMY weather data and simulated internal heat loads. However, when performing Kalman filtering and prediction for the optimal alarm system, these assumptions *will* have to be made. In fact, some of the basic Kalman filtering/prediction mean and covariance propagation equations will change according to these revised assumptions. The counterpart formulae from Chap. 3 that will change are Eqns. 3.9, 3.14, and 3.19 for the predictions. The revised equations are shown in Eqns. 6.19-6.22 below.

$$\hat{\mathbf{q}}_{k+1|k} = \mathbf{A}_d\hat{\mathbf{q}}_{k|k} + \mathbf{B}_d\bar{r} + \mathbf{B}_{w_d}\bar{\mathbf{w}}_k \tag{6.19}$$

$$\mathbf{P}_B^R = \mathbf{A}_d\mathbf{P}_B^R\mathbf{A}_d^T - \mathbf{A}_d\mathbf{P}_B^R\mathbf{C}_d^T(\mathbf{C}_d\mathbf{P}_B^R\mathbf{C}_d^T + \mathbf{R})^{-1}\mathbf{C}_d\mathbf{P}_B^R\mathbf{A}_d^T + \tilde{\mathbf{B}}_d\tilde{\mathbf{W}}\tilde{\mathbf{B}}_d^T \tag{6.20}$$

$$\hat{x}_{k+i|k} \triangleq E[x_{k+i}|x_0,\ldots,x_k] \tag{6.21}$$

$$= \mathbf{C}_d(\mathbf{A}_d^i\hat{\mathbf{q}}_{k|k} + (\mathbf{I}_n - \mathbf{A}_d^i)(\mathbf{I}_n - \mathbf{A}_d)^{-1}(\mathbf{B}_d\bar{r} + \mathbf{B}_{w_d}\bar{\mathbf{w}}_k)) \tag{6.22}$$

where

$$\tilde{\mathbf{W}} = \begin{bmatrix} \sigma^2_{T_{out}} & \mathbf{0}_{1\times m_d} \\ \mathbf{0}_{m_d\times 1} & \mathbf{W} \end{bmatrix}$$

$$\tilde{\mathbf{B}}_d = \begin{bmatrix} \mathbf{B}_d & \mathbf{B}_{w_d} \end{bmatrix}$$

$$\mathbf{P}_B^R = \text{Solution to the DARE (Eqn. 6.20) for the Building Temperature Process}$$

$$\Rightarrow \hat{\mathbf{P}}_B^R = \mathbf{P}_B^R - \mathbf{P}_B^R\mathbf{C}_d^T(\mathbf{C}_d\mathbf{P}_B^R\mathbf{C}_d^T + R)^{-1}\mathbf{C}_d\mathbf{P}_B^R$$

In the design of an optimal alarm system, these formulae are crucial in defining the approximations to the exact alarm regions, or the exact alarm region itself, when computing the Type I/II error probabilities. At the implementation stage, these formulae are necessary to construct the alarm regions based upon the predicted process values, and based upon the optimal P_b selected at the design stage. This allows the alarms to be activated based upon the revised first-principles model, and for a subsequent thermostat setting adjustment.

6.2.4 Revised Cost Function and The Control Law

Now it is of interest to define the cost function for the control objective. Recall that we are now using a cost function that does not include an additive complaint cost term, but rather quantifies the square cost of energy *conditioned* on the type of event being responded to. The cost function can be written as a finite horizon control problem, with window N as follows:

$$J = T_s^2 E[R_N^2 h_N^2 + \sum_{k=0}^{N-1} R_k^2 h_k^2 | \mathcal{E}] \tag{6.23}$$

Our control problem can therefore be posed as a constrained optimization problem for hot events (complaints or alarms), where we want to determine the optimal control sequence $U_0^* = \{u_0^*, u_1^*, \ldots, u_{N-1}^*\}$.

$$\min_{U_0 < 0} J \tag{6.24}$$

And as follows for cold events (complaints or alarms):

$$\min_{U_0 > 0} J \tag{6.25}$$

The reason for the constraints is due to the fact that we would like an *increase* in the thermostat setting (u_k) when there is a cold event, and a *decrease* in the thermostat setting (u_k) when there is a hot event. Therefore, the problem is actually a constrained input optimal LQ control problem, as detailed in Lewis [43]. The quadratic portion of the cost function must be approximated, due to the fact that when h_k^2 is expanded, the cross terms involving p_k and e_k prevent a viable solution to the control problem from being obtained.

As a result, the approximation should be made as follows:

$$\begin{aligned} h_k^2 &= K_p^2(r_k - x_k)^2 + K_p^2 u_k^2 + K_i^2 p_k^2 + 2K_p^2 u_k(r_k - x_k) + 2K_i K_p(r_k + u_k - x_k)p_k \\ &\approx K_p^2(r_k - x_k)^2 + K_p^2 u_k^2 + K_i^2 p_k^2 \end{aligned}$$

Eliminating the last two cross terms involving p_k and e_k has no significant bearing on the cost function or the resulting control law. Only the high frequency components of h_k^2 are affected by eliminating these cross terms. In fact, qualitative evidence of the feasibility of using this approximation is provided by illustrating a plot of h_k^2 vs. its approximation in Figure 6.2.

To find the optimal open-loop control law, we apply the methods of optimal control as defined in Lewis [43]. First we'll rewrite the cost function in Eqn. 6.23 by making the definitions

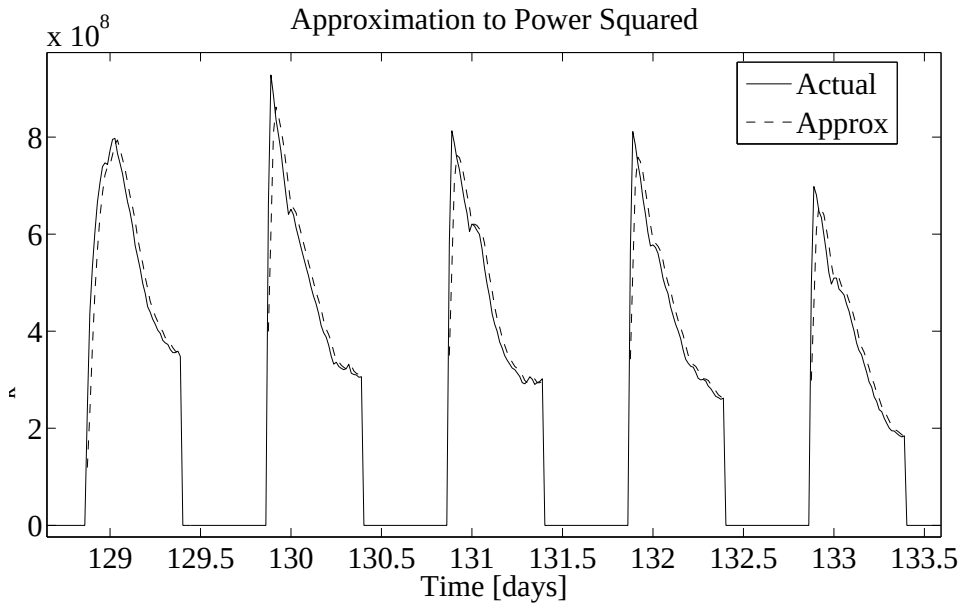

Figure 6.2: h_k^2 and its approximation

below, and using the approximation discussed above.

$$
\begin{aligned}
L(\hat{\mathbf{q}}_{k|\mathcal{E}}, u_k, k) &\triangleq T_s^2 R_k^2 K_p^2 \left[r_k^2 + u_k^2 + \hat{\eta}_{k|\mathcal{E}} - 2\hat{\mathbf{q}}_{k|\mathcal{E}}^T \mathbf{C}_d^T r_k - 2(r_k - \mathbf{C}_d \hat{\mathbf{q}}_{k|\mathcal{E}})\hat{\xi}_{k|\mathcal{E}} + \hat{\mathbf{q}}_{k|\mathcal{E}}^T \mathbf{C}_d^T \mathbf{C}_d \hat{\mathbf{q}}_{k|\mathcal{E}} \right] \\
&+ T_s^2 R_k^2 K_i^2 \left[\hat{\beta}_{k|\mathcal{E}} + 2\hat{\alpha}_{k|\mathcal{E}} \mathbf{M}_d \hat{\mathbf{q}}_{k|\mathcal{E}} + \hat{\mathbf{q}}_{k|\mathcal{E}}^T \mathbf{M}_d^T \mathbf{M}_d \hat{\mathbf{q}}_{k|\mathcal{E}} \right] \\
\phi(\hat{\mathbf{q}}_{N|\mathcal{E}}, N) &\triangleq L(\hat{\mathbf{q}}_{N|\mathcal{E}}, u_N, N) - R_N^2 K_p^2 u_N^2 T_s^2 \\
\Rightarrow J &= \phi(\hat{\mathbf{q}}_{N|\mathcal{E}}, N) + \sum_{k=0}^{N-1} L(\hat{\mathbf{q}}_{k|\mathcal{E}}, u_k, k)
\end{aligned}
$$

Now applying the conditions for optimality according to Lewis [43]:

Recall state equation :

$$\hat{\mathbf{q}}_{k+1|\mathcal{E}} = f(\hat{\mathbf{q}}_{k|\mathcal{E}}, u_k, k)$$

Co-state equation :

$$\boldsymbol{\lambda}_k = \left(\frac{\partial f}{\partial \hat{\mathbf{q}}_{k|\mathcal{E}}}\right)^T \boldsymbol{\lambda}_{k+1} + \frac{\partial L}{\partial \hat{\mathbf{q}}_{k|\mathcal{E}}}$$

Input stationarity condition :

$$0 = \left(\frac{\partial f}{\partial u_k}\right)^T \boldsymbol{\lambda}_{k+1} + \frac{\partial L}{\partial u_k}$$

Boundary condition :

$$0 = \left(\frac{\partial \phi}{\partial \hat{\mathbf{q}}_{N|\mathcal{E}}} - \boldsymbol{\lambda}_N\right)^T d\hat{\mathbf{q}}_{N|\mathcal{E}}$$

Because of the constraints posed in Eqns. 6.24, and 6.25, we must replace the input stationarity condition with the more general Pontryagin's Minimum Principle. Defining the Hamiltonian as follows: $H(\hat{\mathbf{q}}_{k|\mathcal{E}}, u_k, \boldsymbol{\lambda}_{k+1}, k) \triangleq L(\hat{\mathbf{q}}_{k|\mathcal{E}}, u_k, k) + \boldsymbol{\lambda}_{k+1} f(\hat{\mathbf{q}}_{k|\mathcal{E}}, u_k, k)$, we use Pontryagin's Minimum Principle to impose the constraints and deriving the optimal control law, as follows:

$$H(\hat{\mathbf{q}}^*_{k|\mathcal{E}}, u^*_k, \boldsymbol{\lambda}^*_{k+1}, k) \leq H(\hat{\mathbf{q}}^*_{k|\mathcal{E}}, u_k, \boldsymbol{\lambda}^*_{k+1}, k), \forall u_k \in \mathcal{U}$$

The co-state, $\boldsymbol{\lambda}_k$, is necessary to solve the optimal control problem, which often requires the solution to a *two-point boundary value problem*. In doing so, we use the fact the co-state solution can be patterned after the formula: $\boldsymbol{\lambda}_k = \mathbf{P}_k \hat{\mathbf{q}}_{N|\mathcal{E}} - \mathbf{b}_k$, where $\mathbf{b}_k$ and $\mathbf{P}_k$ are auxiliary sequences (both vector and matrix, respectively) to be determined. The set $\mathcal{U}$, is the admissible set, which acts as the constraint in Eqns. 6.24, and 6.25. Therefore, for Eqn. 6.24, the admissible set $\mathcal{U} \equiv U_0 < 0$, and for Eqn. 6.25, the admissible set $\mathcal{U} \equiv U_0 > 0$. When applying the conditions for optimality, we obtain the following dynamic equations necessary to implement the optimal control law:

Kalman Feedforward Gain :

$$\mathbf{K}_k^f = (\mathbf{B}_d^T \mathbf{P}_{k+1} \mathbf{B}_d + 2T_s^2 R_k^2 K_p^2)^{-1} \mathbf{B}_d^T$$

Kalman Feedback Gain :

$$\begin{aligned} \mathbf{K}_k^b &= (\mathbf{B}_d^T \mathbf{P}_{k+1} \mathbf{B}_d + 2T_s^2 R_k^2 K_p^2)^{-1} \mathbf{B}_d^T \mathbf{P}_{k+1} \mathbf{A}_d \\ \Rightarrow \mathbf{K}_k^b &= \mathbf{K}_k^f \mathbf{P}_{k+1} \mathbf{A}_d \end{aligned}$$

Dynamic Riccati Equation :

$$\begin{aligned} \mathbf{P}_k &= \mathbf{A}_d^T \mathbf{P}_{k+1} \mathbf{A}_d - \mathbf{A}_d^T \mathbf{P}_{k+1} \mathbf{B} (2T_s^2 R_k^2 K_p^2 + \mathbf{B}_d^T \mathbf{P}_{k+1} \mathbf{B}_d)^{-1} \mathbf{B}_d^T \mathbf{P}_{k+1} \mathbf{A}_d \\ &\quad + 2T_s^2 R_k^2 (K_p^2 \mathbf{C}_d^T \mathbf{C}_d + K_i^2 \mathbf{M}_d^T \mathbf{M}_d) \\ \Rightarrow \mathbf{P}_k &= \mathbf{A}_d^T \mathbf{P}_{k+1} (\mathbf{A}_d - \mathbf{B}_d \mathbf{K}_k^b) + 2T_s^2 R_k^2 (K_p^2 \mathbf{C}_d^T \mathbf{C}_d + K_i^2 \mathbf{M}_d^T \mathbf{M}_d) \end{aligned}$$

Boundary Condition :

$$\mathbf{P}_N = 2T_s^2 R_N^2 (K_p^2 \mathbf{C}_d^T \mathbf{C}_d + K_i^2 \mathbf{M}_d^T \mathbf{M}_d)$$

Auxiliary Sequence :

$$\begin{aligned} \mathbf{b}_k &= (\mathbf{A}_d - \mathbf{B}_d \mathbf{K}_k^b)^T (\mathbf{b}_{k+1} - \mathbf{P}_{k+1} \mathbf{B}_{w_d} \bar{\mathbf{w}}_k) \\ &\quad + (\mathbf{G}_k^\xi - (\mathbf{K}_k^b)^T) r_k - \mathbf{G}_k^\xi \hat{\xi}_{k|\mathcal{E}} - \mathbf{G}_k^\alpha \hat{\alpha}_{k|\mathcal{E}} \end{aligned}$$

Boundary Condition :

$$\mathbf{b}_N = 2T_s^2 R_N^2 \left[K_p^2 \mathbf{C}_d^T (\hat{\xi}_{N|\mathcal{E}} - r_N) + K_i^2 \mathbf{M}_d^T \hat{\alpha}_{N|\mathcal{E}} \right]$$

where

$$\begin{aligned} \mathbf{G}_k^\xi &= 2T_s^2 R_k^2 K_p^2 \mathbf{C}_d^T \\ \mathbf{G}_k^\alpha &= 2T_s^2 R_k^2 K_i^2 \mathbf{M}_d^T \end{aligned}$$

In solving for the boundary conditions, we assume a free final state situation, therefore,

$d\hat{\mathbf{q}}_{N|\mathcal{E}} \neq 0$, and $\boldsymbol{\lambda}_N = \frac{\partial \phi}{\partial \hat{\mathbf{q}}_{N|\mathcal{E}}}$. When applying Pontryagin's Minimum Principle, the optimal control law can be summarized in Eqns. 6.26-6.29.

$$\text{Let } d_{k+1} \triangleq -\frac{1}{2T_s^2 R_k^2 K_p^2}\mathbf{B}_d^T \boldsymbol{\lambda}_{k+1} \tag{6.26}$$

$$= -\mathbf{K}_k^b \hat{\mathbf{q}}_{k|\mathcal{E}} + \mathbf{K}_k^f \left[\mathbf{b}_{k+1} - \mathbf{P}_{k+1}(\mathbf{B}_d \bar{r} + \mathbf{B}_{w_d}\bar{\mathbf{w}}_k)\right] \tag{6.27}$$

$$\text{Let } s(\delta) \triangleq \begin{cases} 1 & \delta > 0 \\ 0 & \delta \leq 0 \end{cases}$$

$\Rightarrow$ *For Cold Events :*

$$u_k = d_{k+1} s(d_{k+1}) \tag{6.28}$$

$\Rightarrow$ *For Hot Events :*

$$u_k = d_{k+1} s(-d_{k+1}) \tag{6.29}$$

Therefore, the optimal control is actually nonlinear due to the step function defined as $s(\delta)$. At the time of the event, we choose $\hat{\mathbf{q}}_{k|\mathcal{E}} \equiv \hat{\mathbf{q}}_{k|k}$, or the optimal state estimate as the initial state when $k = 0$. Furthermore, we use Eqn. 6.14 to step the optimal control law forward in time from $k = 0$ to $k = N$. Note that the use of $k = 0$ here is in poor form, and it is actually a generalization of $k = k_{\mathcal{E}}$, or the time at which a complaint or alarm event occurs.

6.2.5 Conditioning on Complaint and Alarm Information

What makes this optimal control law different than a standard optimal open-loop LQ tracking problem is the fact that we've conditioned on hot or cold events. The $\hat{\alpha}_{k|\mathcal{E}}$ and $\hat{\xi}_{k|\mathcal{E}}$ terms of the auxiliary sequence $\mathbf{b}_k$ add to the uniqueness of our optimal control law. The origin of these terms can be attributed to conditioning on complaint or alarm information. We use the information dynamically in order to integrate the state information into the control response at the time of the event, based upon the relative complaint or alarm statistics. The "relative" statistics characterize the mean and variance of the *difference* between the building temperature and the hot or cold complaint

levels. As such, this final subsection details the dynamic computation of these two terms which use the relative statistics, for both complaints and alarms. Complaints will be categorized into two distinct types: arrival complaints only, and operating complaints only. Alarms will be categorized into two types as well, defined previously in Chap. 3: both arrival and operating complaints, and operating complaints only.

Recall the following definitions:

$$\hat{\alpha}_{k|\mathcal{E}} \triangleq E[\tilde{p}_k|\mathcal{E}]$$
$$\hat{\xi}_{k|\mathcal{E}} \triangleq E[\tilde{x}_k|\mathcal{E}]$$

Now we'd like to expand these definitions to be more specific as to the type of event being responded to. Furthermore, the indexing needs to be augmented so that the equations are more compatible for use with the control law shown in Sec. 6.2.4. Let's first consider the case of response to complaints. When considering operating complaints, they can most easily be described as zero-level upcrossings of the process z_k. Recall the definition of z_k from Chap. 3: $z_k \triangleq x_k - y_k$ when considering the interaction between the hot complaint process (y_k) and the building temperature process (x_k), and $z_k \triangleq y_k - x_k$ when considering the interaction between the cold complaint process (y_k) and the building temperature process (x_k). Therefore, in general for operating complaints (C_o), Eqns. 6.30-6.35 hold true.

$$\hat{\xi}_{k+i|C_o} \triangleq E[\tilde{x}_{k+i}|C_o] \tag{6.30}$$

$$= E[\tilde{x}_{k+i}|z_{k-1}<0, z_k>0] \tag{6.31}$$

$$= \frac{\int_0^\infty \int_{-\infty}^0 E[\tilde{x}_{k+i}|\mathbf{x}]\mathcal{N}(\mathbf{x};\mu_\mathbf{x},\Sigma_\mathbf{x})d\mathbf{x}}{\int_0^\infty \int_{-\infty}^0 \mathcal{N}(\mathbf{x};\mu_\mathbf{x},\Sigma_\mathbf{x})d\mathbf{x}} \tag{6.32}$$

$$\hat{\alpha}_{k+i|C_o} \triangleq E[\tilde{p}_{k+i}|C_o] \tag{6.33}$$

$$= E[\tilde{p}_{k+i}|z_{k-1}<0, z_k>0] \tag{6.34}$$

$$= \frac{\int_0^\infty \int_{-\infty}^0 E[\tilde{p}_{k+i}|\mathbf{x}]\mathcal{N}(\mathbf{x};\mu_\mathbf{x},\Sigma_\mathbf{x})d\mathbf{x}}{\int_0^\infty \int_{-\infty}^0 \mathcal{N}(\mathbf{x};\mu_\mathbf{x},\Sigma_\mathbf{x})d\mathbf{x}} \tag{6.35}$$

$$\text{where } \mathbf{x} = \begin{bmatrix} z_{k-1} \\ z_k \end{bmatrix}$$

$$\mu_\mathbf{x} = \begin{cases} \begin{bmatrix} \mathbf{C}_d\mathbf{q}_{ss} - \mu_y \\ \mathbf{C}_d\mathbf{q}_{ss} - \mu_y \end{bmatrix} & \text{hot complaint} \\ \begin{bmatrix} \mu_y - \mathbf{C}_d\mathbf{q}_{ss} \\ \mu_y - \mathbf{C}_d\mathbf{q}_{ss} \end{bmatrix} & \text{cold complaint} \end{cases}$$

$$\Sigma_\mathbf{x} = \begin{bmatrix} \mathbf{C}_d\mathbf{P}_B^L\mathbf{C}_d^T + \mathbf{C}\mathbf{P}_{ss}^L\mathbf{C}^T + R & \mathbf{C}_d\mathbf{P}_B^L\mathbf{A}_d^T\mathbf{C}_d^T + \mathbf{C}\mathbf{P}_{ss}^L\mathbf{A}^T\mathbf{C}^T \\ \mathbf{C}_d\mathbf{A}_d\mathbf{P}_B^L\mathbf{C}_d^T + \mathbf{C}\mathbf{A}\mathbf{P}_{ss}^L\mathbf{C}^T & \mathbf{C}_d\mathbf{P}_B^L\mathbf{C}_d^T + \mathbf{C}\mathbf{P}_{ss}^L\mathbf{C}^T + R \end{bmatrix}$$

$$E[\tilde{x}_{k+i}|\mathbf{x}] = \mathbf{C}_d\mathbf{P}_B^L \begin{bmatrix} \mathbf{C}_d\mathbf{A}_d^{i+1} \\ \mathbf{C}_d\mathbf{A}_d^{i} \end{bmatrix}^T \Sigma_\mathbf{x}^{-1}(\mathbf{x}-\mu_\mathbf{x})$$

$$E[\tilde{p}_{k+i}|\mathbf{x}] = \mathbf{C}_d\mathbf{P}_B^L \begin{bmatrix} \mathbf{M}_d\mathbf{A}_d^{i+1} \\ \mathbf{M}_d\mathbf{A}_d^{i} \end{bmatrix}^T \Sigma_\mathbf{x}^{-1}(\mathbf{x}-\mu_\mathbf{x})$$

And $\mathbf{P}_B^L$ represents the solution to the Lyapunov equation for the closed-loop building temperature process model, such that: $\mathbf{P}_B^L = \mathbf{A}_d\mathbf{P}_B^L\mathbf{A}_d^T + \tilde{\mathbf{B}}_d\tilde{\mathbf{W}}\tilde{\mathbf{B}}_d^T$. Furthermore, μ_y, $\mathbf{C}$, $\mathbf{A}$, and $\mathbf{P}_{ss}^L$, are all generic expressions for the mean, state-space matrices, and unconditional algebraic state covariance for the complaint level processes, respectively. The variable $\mathbf{q}_{ss}$ represents the

steady-state of the building temperature process that is computed using Eqn. 6.5. The last term involving v_k is neglected when computing the steady-state $\mathbf{q}_{ss}$, since $E[v_k] = 0$ due to our zero-mean Gaussian measurement noise assumption of $v_k \sim \mathcal{N}(0, R)$. When considering arrival complaints (C_a), we'll describe them as two consecutive exceedances at any point during the arrival period, or as $z_{k-1} > 0, z_k > 0$. Therefore, the exact same equations apply as shown in Eqns. 6.30-6.35, with minor modifications on the limits of integration as shown below in Eqns. 6.36-6.41.

$$\hat{\xi}_{k+i|C_a} \triangleq E[\tilde{x}_{k+i}|C_a] \tag{6.36}$$

$$= E[\tilde{x}_{k+i}|z_{k-1} > 0, z_k > 0] \tag{6.37}$$

$$= \frac{\int_0^\infty \int_0^\infty E[\tilde{x}_{k+i}|\mathbf{x}]\mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})d\mathbf{x}}{\int_0^\infty \int_0^\infty \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})d\mathbf{x}} \tag{6.38}$$

$$\hat{\alpha}_{k+i|C_a} \triangleq E[\tilde{p}_{k+i}|C_a] \tag{6.39}$$

$$= E[\tilde{p}_{k+i}|z_{k-1} > 0, z_k > 0] \tag{6.40}$$

$$= \frac{\int_0^\infty \int_0^\infty E[\tilde{p}_{k+i}|\mathbf{x}]\mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})d\mathbf{x}}{\int_0^\infty \int_0^\infty \mathcal{N}(\mathbf{x}; \mu_\mathbf{x}, \Sigma_\mathbf{x})d\mathbf{x}} \tag{6.41}$$

Now let's consider the case of response to alarms. The exact alarm region cannot easily be integrated over, and therefore we should use the most accurate approximation available to us. As such, the most correct treatment for alarms during the "operating complaint only" period should be with the multi-dimensional approximation. Computational efficiency is not an issue due to the fact that the design stage is assumed to be complete, and for control we are concerned only with implementation based upon the optimally selected P_b value. Recall from Chap. 3 the approximation to the alarm region for "operating complaints only" in Eqn. 3.53:

$$
\begin{aligned}
A_{approx} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{approx}} \subset \mathbb{R}^{m+1}\} \\
&= \{\hat{\mathbb{X}} : A_0 \bigcap \left[\bigcup_{i=1}^{m} A_i\right]\} \\
\text{where } A_0 &\triangleq \hat{x}_{k|k} \leq \overbrace{L - \sqrt{V_{k|k}}\Phi^{-1}(P_b)}^{\mathcal{X}_0} \\
A_i &\triangleq \hat{x}_{k+i|k} \geq \underbrace{L + \sqrt{V_{k+i|k}}\Phi^{-1}(P_b)}_{\mathcal{Y}_i} \\
\forall i &\in 1, \ldots, m \\
\text{Or, } A_{approx} &= \{\hat{\mathbb{X}} : \hat{x}_{k|k} \leq L - \sqrt{V_{k|k}}\Phi^{-1}(P_b) \bigcap \left[\bigcup_{i=1}^{m} \hat{x}_{k+i|k} \geq L + \sqrt{V_{k+i|k}}\Phi^{-1}(P_b)\right]\}
\end{aligned}
$$

Note that in this case, $L = 0$. As defined, multi-dimensional approximation of the exact alarm region spans all of the time slices in the horizon. Therefore, its use for implementation in the control law presented in Sec. 6.2.4 may not be feasible on a step-by-step basis. Consequently, instead of treating the alarm region as a whole, we will split it up into pieces. The desired quantities necessary for use in the control law for operating alarms only are shown in Eqns. 6.42-6.43.

$$\hat{\xi}_{k+i|A_o} \triangleq E[\tilde{x}_{k+i}|A_o] \tag{6.42}$$

$$\hat{\alpha}_{k+i|A_o} \triangleq E[\tilde{p}_{k+i}|A_o] \tag{6.43}$$

To split the alarm region into pieces, we will consider each "step," i.e. $\hat{\xi}_{k+i|A_o}$ or $\hat{\alpha}_{k+i|A_o}$ to be associated with the i^{th} step from the initial time (that the control algorithm starts at). This i^{th} step will correspond to the $A_0 \bigcap A_i$ portion of the alarm region: $A_0 \bigcap [\bigcup_{i=1}^{m} A_i]$. When $i = 0$, then it reduces to just A_0. We now summarize and expand on Eqns. 6.42-6.43 with Eqns. 6.44-6.49:

$$\hat{\xi}_{k+i|A_o} \stackrel{\triangle}{=} E[\tilde{x}_{k+i}|A_o] \tag{6.44}$$

$$= \begin{cases} E[\tilde{x}_{k+i}|A_0] & i=0 \\ E[\tilde{x}_{k+i}|A_0, A_i] & i>0 \end{cases} \tag{6.45}$$

$$= \begin{cases} E[\tilde{x}_{k+i}|\hat{z}_{k|k} < \mathcal{X}_0] & i=0 \\ E[\tilde{x}_{k+i}|\hat{z}_{k|k} < \mathcal{X}_0, \hat{z}_{k+i|k} > \mathcal{Y}_i] & i>0 \end{cases} \tag{6.46}$$

$$= \begin{cases} \frac{\int_{-\infty}^{\mathcal{X}_0} E[\tilde{x}_k|\hat{z}_{k|k}]\mathcal{N}\left(\hat{z}_{k|k};\mu_{\hat{z}_{k|k}},\mathbf{C}_d(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R)\mathbf{C}_d^T\right)d\hat{z}_{k|k}}{\int_{-\infty}^{\mathcal{X}_0}\mathcal{N}\left(\hat{z}_{k|k};\mu_{\hat{z}_{k|k}},\mathbf{C}_d(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R)\mathbf{C}_d^T\right)d\hat{z}_{k|k}} & i=0 \\ \frac{\int_{-\infty}^{\mathcal{X}_0}\int_{\mathcal{Y}_i}^{\infty} E[\tilde{x}_{k+i}|\mathbf{x}]\mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})d\mathbf{x}}{\int_{-\infty}^{\mathcal{X}_0}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})d\mathbf{x}} & i>0 \end{cases} \tag{6.47}$$

$$\hat{\alpha}_{k+i|A_o} \stackrel{\triangle}{=} E[\tilde{p}_{k+i}|A_o] \tag{6.48}$$

$$= \begin{cases} \frac{\int_{-\infty}^{\mathcal{X}_0} E[\tilde{p}_k|\hat{z}_{k|k}]\mathcal{N}\left(\hat{z}_{k|k};\mu_{\hat{z}_{k|k}},\mathbf{C}_d(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R)\mathbf{C}_d^T\right)d\hat{z}_{k|k}}{\int_{-\infty}^{\mathcal{X}_0}\mathcal{N}\left(\hat{z}_{k|k};\mu_{\hat{z}_{k|k}},\mathbf{C}_d(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R)\mathbf{C}_d^T\right)d\hat{z}_{k|k}} & i=0 \\ \frac{\int_{-\infty}^{\mathcal{X}_0}\int_{\mathcal{Y}_i}^{\infty} E[\tilde{p}_{k+i}|\mathbf{x}]\mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})d\mathbf{x}}{\int_{-\infty}^{\mathcal{X}_0}\int_{\mathcal{Y}_i}^{\infty}\mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})d\mathbf{x}} & i>0 \end{cases} \tag{6.49}$$

$$
\begin{aligned}
\text{where } E[\tilde{x}_k|\hat{z}_{k|k}] &= \begin{cases} \hat{z}_{k|k} - (\mathbf{C}_d\mathbf{q}_{ss} - \mu_y) & \text{hot alarm} \\ \hat{z}_{k|k} - (\mu_y - \mathbf{C}_d\mathbf{q}_{ss}) & \text{cold alarm} \end{cases} \\
E[\tilde{p}_k|\hat{z}_{k|k}] &= \begin{cases} \frac{\mathbf{M}_d(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R)\mathbf{C}_d^T}{\mathbf{C}_d(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R)\mathbf{C}_d^T}(\hat{z}_{k|k} - (\mathbf{C}_d\mathbf{q}_{ss} - \mu_y)) & \text{hot alarm} \\ \frac{\mathbf{M}_d(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R)\mathbf{C}_d^T}{\mathbf{C}_d(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R)\mathbf{C}_d^T}(\hat{z}_{k|k} - (\mu_y - \mathbf{C}_d\mathbf{q}_{ss})) & \text{cold alarm} \end{cases} \\
\text{and } \mu_{\hat{z}_{k|k}} &= \mu_z = \begin{cases} \mathbf{C}_d\mathbf{q}_{ss} - \mu_y & \text{hot alarm} \\ \mu_y - \mathbf{C}_d\mathbf{q}_{ss} & \text{cold alarm} \end{cases} \\
\mathbf{x} &= \begin{bmatrix} \hat{z}_{k|k} \\ \hat{z}_{k+i|k} \end{bmatrix} \\
\mu_{\mathbf{x}} &= \begin{cases} \begin{bmatrix} \mathbf{C}_d\mathbf{q}_{ss} - \mu_y \\ \mathbf{C}_d\mathbf{q}_{ss} - \mu_y \end{bmatrix} & \text{hot alarm} \\ \begin{bmatrix} \mu_y - \mathbf{C}_d\mathbf{q}_{ss} \\ \mu_y - \mathbf{C}_d\mathbf{q}_{ss} \end{bmatrix} & \text{cold alarm} \end{cases} \\
\Sigma_{\mathbf{x}} &= \begin{bmatrix} \mathbf{C}_d \\ \mathbf{C}_d\mathbf{A}_d^i \end{bmatrix} (\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R) \begin{bmatrix} \mathbf{C}_d \\ \mathbf{C}_d\mathbf{A}_d^i \end{bmatrix}^T \\
E[\tilde{x}_{k+i}|\mathbf{x}] &= \mathbf{C}_d\mathbf{A}_d^i(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R) \begin{bmatrix} \mathbf{C}_d \\ \mathbf{C}_d\mathbf{A}_d^i \end{bmatrix}^T \Sigma_{\mathbf{x}}^{-1}(\mathbf{x} - \mu_{\mathbf{x}}) \\
E[\tilde{p}_{k+i}|\mathbf{x}] &= \mathbf{M}_d\mathbf{A}_d^i(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R) \begin{bmatrix} \mathbf{C}_d \\ \mathbf{C}_d\mathbf{A}_d^i \end{bmatrix}^T \Sigma_{\mathbf{x}}^{-1}(\mathbf{x} - \mu_{\mathbf{x}})
\end{aligned}
$$

Finally, for alarms characterizing "both the arrival and the operating complaint" period, we again use the multi-dimensional approximation. Therefore, recall from Chap. 3 the approximation to the alarm region for both arrival and operating complaints:

$$
\begin{aligned}
A_{approx} &\triangleq \{\hat{\mathbb{X}} \in \Omega_{A_{approx}} \subset \mathbb{R}^{m+1}\} \\
&= \{\hat{\mathbb{X}} : \bigcup_{i=0}^{m} A_i\} \\
&= \{\hat{\mathbb{X}} : \bigcup_{i=0}^{m} \hat{x}_{k+d+i|k} \geq \underbrace{L + \sqrt{V_{k+d+i|k}}\Phi^{-1}(P_b)}_{\mathcal{Y}_i}\}
\end{aligned}
$$

Note again that in this case, $L = 0$, and unlike the case of operating complaints only, there is no distinction between $i = 0$ and $i > 0$. To close out this section, the final desired terms of the control law for both arrival and operating alarms only are shown in Eqns. 6.50-6.53.

$$
\begin{aligned}
\hat{\xi}_{k+i|A_a} &\triangleq E[\tilde{x}_{k+i}|A_a] && (6.50)\\
&= \frac{\int_{\mathcal{Y}_i}^{\infty} E[\tilde{x}_{k+i}|\mathbf{x}]\mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})d\mathbf{x}}{\int_{\mathcal{Y}_i}^{\infty} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})d\mathbf{x}} && (6.51)\\
E[\tilde{x}_{k+i}|\mathbf{x}] &= \mathbf{C}_d\mathbf{A}_d^i(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R)(\mathbf{A}_d^T)^{d+i}\mathbf{C}_d^T\Sigma_{\mathbf{x}}^{-1}(\mathbf{x} - \mu_{\mathbf{x}})
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{\alpha}_{k+i|A_a} &\triangleq E[\tilde{p}_{k+i}|A_a] && (6.52)\\
&= \frac{\int_{\mathcal{Y}_i}^{\infty} E[\tilde{p}_{k+i}|\mathbf{x}]\mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})d\mathbf{x}}{\int_{\mathcal{Y}_i}^{\infty} \mathcal{N}(\mathbf{x};\mu_{\mathbf{x}},\Sigma_{\mathbf{x}})d\mathbf{x}} && (6.53)\\
E[\tilde{p}_{k+i}|\mathbf{x}] &= \mathbf{M}_d\mathbf{A}_d^i(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R)(\mathbf{A}_d^T)^{d+i}\mathbf{C}_d^T\Sigma_{\mathbf{x}}^{-1}(\mathbf{x} - \mu_{\mathbf{x}})
\end{aligned}
$$

$$
\begin{aligned}
\text{where } \mathbf{x} &= \hat{z}_{k+i+d|k} \\
\mu_{\mathbf{x}} &= \begin{cases} \mathbf{C}_d\mathbf{q}_{ss} - \mu_y & \text{hot alarm} \\ \mu_y - \mathbf{C}_d\mathbf{q}_{ss} & \text{cold alarm} \end{cases} \\
\Sigma_{\mathbf{x}} &= \mathbf{C}_d\mathbf{A}_d^{d+i}(\mathbf{P}_B^L - \hat{\mathbf{P}}_B^R)(\mathbf{A}_d^T)^{d+i}\mathbf{C}_d^T
\end{aligned}
$$

6.3 Discussion

6.3.1 Computational Requirements

The integrals in Sec. 6.2.5 are not as computationally intensive as those that are required during the optimal alarm design stage, which have a much higher dimension in general. This is advantageous because the optimal control algorithm is executed on-line in real-time. However, the same routine modified from Genz's code [27] is used to compute them. Traditionally, this code is more effective for higher-dimensional integrations, but can be used just as well for lower dimensional ones. Most of the integrals from Sec. 6.2.5 are either univariate or bivariate Gaussian -based expectation or probability computations. The univariate probabilities may be computed more basically with the $\Phi(\cdot)$ cumulative normal distribution function. The numerators of the univariate expectation computations may even be found analytically.

Another computational issue in implementing the control law is its time-varying nature. Specifically, the price of energy, R_k, is included in the control law, so the optimal control law is demand responsive as well. This goes for any time-of-use energy scenario where the price of energy can be predicted for at least a short finite window into the future. For most time-of-use energy rate schemes, they vary according to a fixed schedule, based on the time of day and the season. The rate scenario to be used for simulations in the subsequent chapter is the E-19 electrical energy tariff offered by PG&E, as used by Federspiel in his study [21, 22]. However, as long as the price is known over a finite prediction horizon, the control law is well-suited for real-time pricing scenarios as well.

Of the parameters used in the control law, the finite horizon to be used is of great importance. Recall that the time step used for optimal alarm deemed to be most suitable in Chap. 3 was $T_s = 20$ min. A 2-hour window is sufficient time for the control or alarm response by thermostat setting modification to be detected by building occupants, according to the dynamics of the building temperature process. As such, we will use 6 steps for the horizon for both complaints and alarms. Furthermore, it is a reasonable time window when considering the critical "arrival" and "operating"

time periods under consideration. Although the optimal alarm results presented in Chap. 3 used an $m = 4$-hour period, the look-ahead arrival period of $d = 1$-hour remains unchanged. The discrepancy between the 4-hour period and the 2-hour period is not significant in terms of the issues mentioned above. Rather, it allows for a reduced computational burden when implementing the control algorithm.

6.3.2 Preliminary Results

Recall that the control law presented in Sec. 6.2.4 was based upon Eqn. 6.14, whose purpose was to step the optimal control law forward in time from $k = 0$ to $k = N$. Also, at the time of the event, we chose $\hat{\mathbf{q}}_{k|\mathcal{E}} \equiv \hat{\mathbf{q}}_{k|k}$, or the optimal state estimate as the initial state when $k = 0$. Therefore, because the control uses an estimate of the state of the plant at the time of execution, the control law will not be identical for all events. However, it is of use to obtain an idea of the average control response for both complaint and alarm events. As such, instead of using $\hat{\mathbf{q}}_{k|\mathcal{E}} \equiv \hat{\mathbf{q}}_{k|k}$, we will study the expected value of $\hat{\mathbf{q}}_{k|k}$, which is $E[\hat{\mathbf{q}}_{k|k}] = \mathbf{q}_{ss}$. The steady-state is fixed, and also acts as the initial state for simulations. We can now use this as the baseline for the average control response to any type of event.

It is also of interest to investigate the effect of different modelling choices when implementing the control law. Specifically, there are different ways to treat the mean of the input noise when implementing the control algorithm when using Eqn. 6.14. The equation can be stepped forward in time with disturbance previews using the following methods:

1. Use the actual stochastic input disturbance, $\mathbf{w}_k$.
2. Use stochastic mean, or $\bar{\mathbf{w}}_k$ by implementing the EWMA method for outside temperature and expected internal heat gain as defined in Sec. 6.2.3.
3. Use a pseudo-deterministic seasonal average for $\bar{\mathbf{w}}_k$ (therefore this changes 4 times per year).
4. Use a deterministic annual average for $\bar{\mathbf{w}}_k$.

5. Use a deterministic average of $\bar{\mathbf{w}}_k = \mathbf{0}_{m_d \times 1}$.

The control and alarm responses for all of the above cases are shown in Fig. 6.3 (hot) and Fig. 6.4 (cold). The optimal control, u_k, is shown in both figures, over the desired finite horizon, and it represents the deviation from the nominal thermostat setting, $\bar{r}$. For completeness, thick solid lines are also shown in all of the illustrations, representing the optimal setpoint modification policy *without* the benefit of conditioning on complaint or alarm information. In other words, the cost function degenerates to penalizing energy only, so the optimal strategy is to change the setpoint according to a zero energy output, balanced against all of the other influences of energy in the model. The input disturbances influence this zero-energy usage setpoint policy as well as the building thermal mass. The simulation start time for all of the illustrations shown is Feb. 23, where the TMY weather data is for Seattle, Washington. Clearly, the model of building thermal mass is sufficient to balance against the input disturbances, in fact, it overcompensates for the outside temperature. This is due to the fact the zero-energy usage setpoint policy is always negative, indicating a need to decrease the temperature in order to zero out the need for any actuator energy, even in winter.

When we consider conditioning on complaint or alarm information, the remaining non-solid lines (including the cross 'x' sequence) in all of the illustrations represent the optimal setpoint modification policy for Cases # 1-4 listed above. Clearly, due to the fact that they are all fairly well superimposed, there is no significant difference among any of these modelling nuances. Case # 5, shown as a solid line of normal thickness is the only slightly different situation. Here it is assumed that $\bar{\mathbf{w}}_k = \mathbf{0}_{m_d \times 1}$, which is the least desirable choice. An important fact to consider for *all* cases, excluding the zero-energy case, is that the price of energy during the response period has a slight change during the course of the response. This change, shown in Fig. 6.5, will significantly impact the qualitative nature of the control response to events. Recall that the simulation start time for the period shown in Fig. 6.5 is Feb. 23, therefore, the seasonal rates for winter are in effect.

For the cases in which we are considering response to alarms as opposed to events, the control strategy is also heavily dependent upon the optimally selected P_b value. This concept will

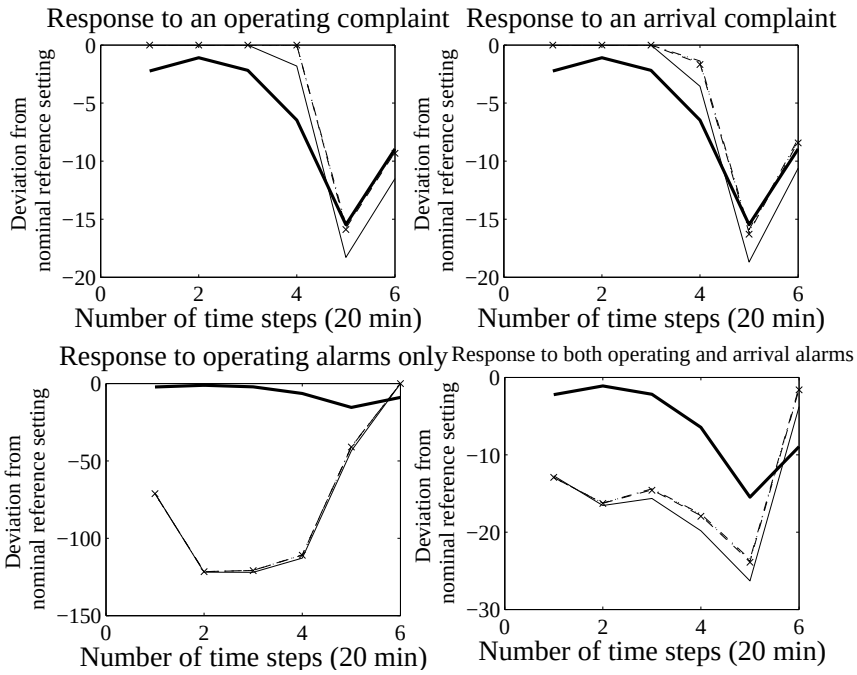

Figure 6.3: Average Responses to Hot Events and Comparison of Differing Modelling Choices

be presented in earnest in Chap. 7. For the alarm responses shown in Figs. 6.3 and 6.4, the optimal P_b values are shown in Table 6.1 below. They have been chosen according to the minimax design criterion, using the sub-interval method, and modelled with downcrossings when considering arrivals.

Hot Arrival & Operating	0.0042572
Cold Arrival & Operating	0.027867
Hot Operating	0.0053595
Cold Operating	0.036319

Table 6.1: Optimal Alarm System Design P_b Values

The extreme responses to the optimal alarm system shown illustrate the sensitivity of the control algorithm to the value of P_b. Had P_b been selected otherwise, or according to some other criteria, the response would have been significantly different. There are actually two extremes. One extreme is shown for operating complaints only, on the lower left corner of both hot and cold responses in Figs. 6.3 and 6.4. With the other extreme, we see for both cold operating and arrival

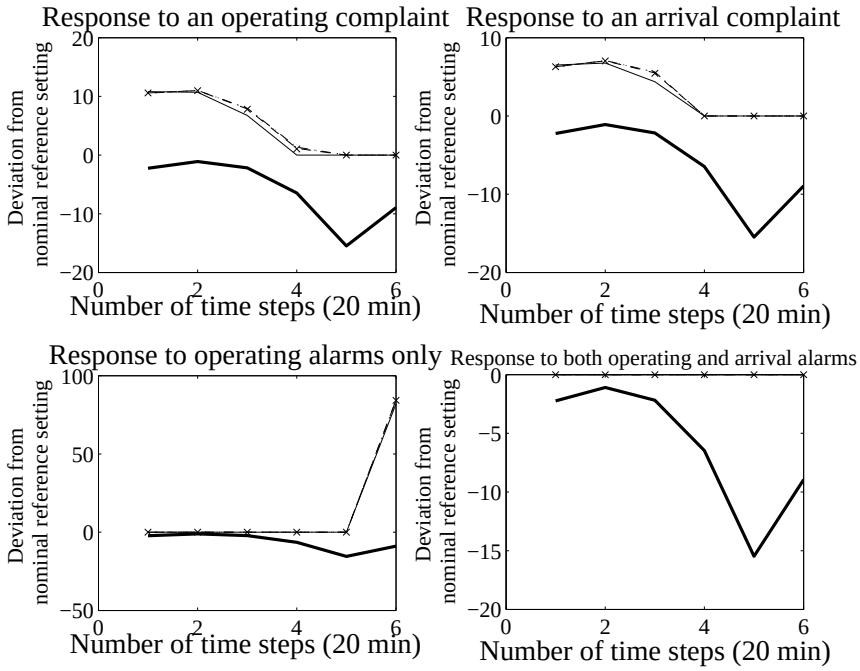

Figure 6.4: Average Responses to Cold Events and Comparison of Differing Modelling Choices

alarms, on the lower right hand side of Fig. 6.4, there is no response (note the cross 'x' sequence). This is due to the relatively low probability of alarm (not shown) for the corresponding value of P_b as compared to the others in shown in Table 6.1. Therefore the respective optimal alarm system will have a much higher sensitivity. False alarms will be quite frequent, therefore the optimal control policy will be to temper the response with little action. This action is taken with the appropriate cold constraint imposed. The next chapter will go into more depth concerning sensitivity issues with regard to P_b, the different selection criteria for optimal alarm design, and practical implementation for both alarm and control.

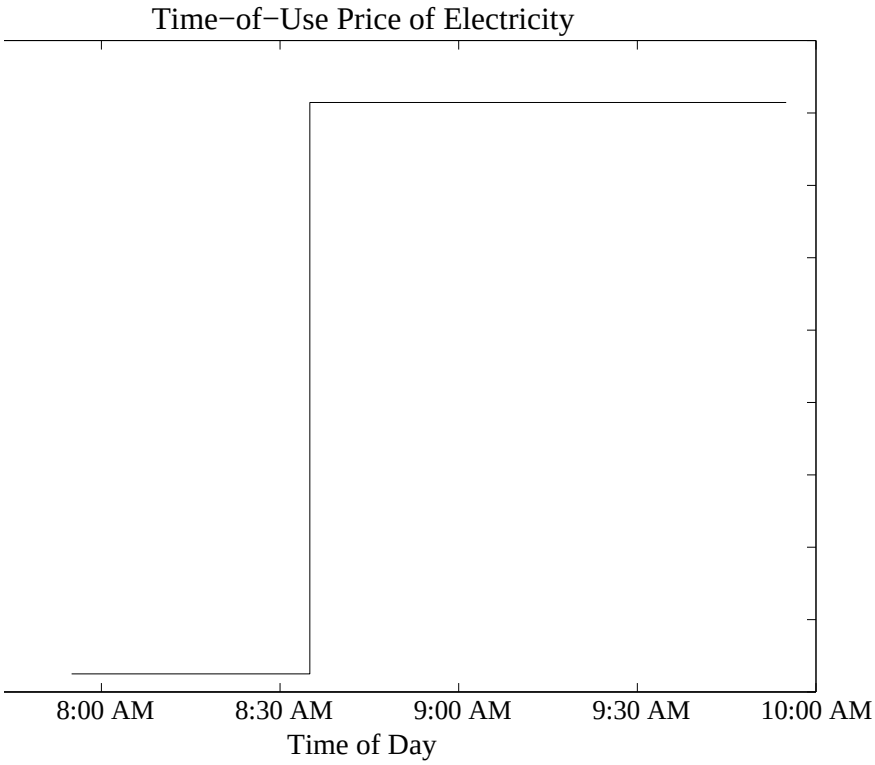

Figure 6.5: Price of Electricity (Rate) During Simulation Period

Chapter 7

Practical Implementation

7.1 Selecting a Prototype Building/Zone

Having a prototype building and zone to use as a testbed for simulations is of great importance. It will allow us to study the utility of the optimal alarm and control algorithms for important metrics under a common framework. Suitable candidates for the prototype include the buildings that were chosen to re-calibrate the complaint prediction model in Federspiel's study [21, 22], giving us the parameters in Table 2.3. These buildings were described in depth in the study, and the subset that we will deem as candidates here will come from the same geographic location (Seattle, WA), for ease of analysis when using TMY data in simulations. These buildings are denoted as Buildings A, C, and F. Building A is a small 60,000 square foot laboratory/office building. Building C is medium-sized 284,000 square foot converted industrial facility that serves primarily as office space. Building F is a 798,000 square foot large hi-rise commercial office building in a downtown area. The buildings all have commercial maintenance management systems that store relevant data, and the number of zones to choose from is proportional to the square footage.

In selecting a prototype among the various buildings and zones, recall that the design of an optimal control algorithm required a model based upon first principles. As such, an important

criterion is that we must find a building and zone whose temperature dynamics have at the very least a qualitative similarity to the model we've used as the basis for optimal control. Of course, this model will adhere to the same scheduling requirements as would normally be in place for the real building/zone. In consideration of this, it was found that Building C was the most suitable candidate among the three. Qualitatively, the model's building temperature appears to have reasonable similarity to a particular zone within the building, as illustrated in the week-long simulation period shown in Fig. 7.1.

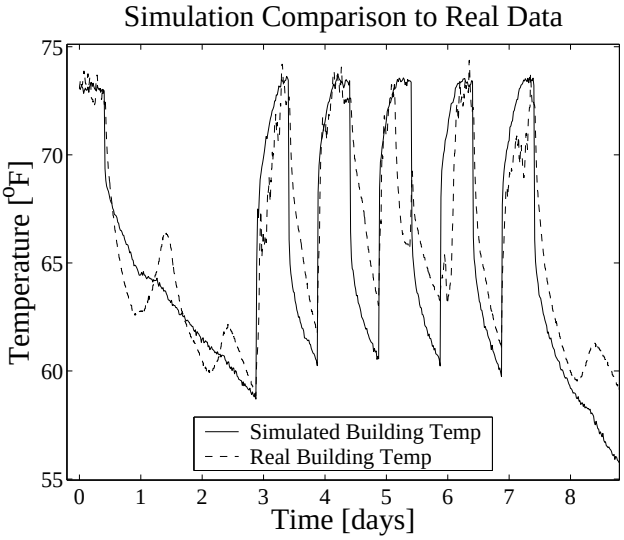

Figure 7.1: Real vs. Modelled Building Temperature Time-Series Data

7.2 Optimal Alarm Results

7.2.1 Selecting the Optimal P_b Value and Alarm System Utility

The results presented in Chaps. 3 and 5 did not provide conclusive results on how an actual optimal alarm system is designed and implemented in practice. This section is devoted to providing closure on some of these issues. Recall the conclusion from Chap. 3 highlighting the importance of

designing and implementing an optimal alarm system by selecting the optimal P_b value.

In Chap. 3, the assumed rule of thumb was that we are probably much better off missing a possible customer complaint and saving the money in labor hours and dollars charged to respond, than anticipating and responding to a complaint that will never happen. Therefore, it was deemed that false alarms have a higher cost in our particular application. There were many cost functions and other alternatives for selecting the optimal value of P_b also discussed in Chap. 3. In fact, a new method for minimizing false alarms was introduced, by reformulating the penalty further to maximizing the expected number of complaints conditioned on an alarm, $E[C|A]$. The details and computational advantages of using this revised cost function were presented in Chap. 5.

To illustrate some of the issues with using this technique, we use the "Federspiel model" from Chap. 3 as an example. It characterizes the building and complaint process interaction that is of greatest interest to the thesis objective. The parameters described in Chap. 3 apply for the case of operating complaints only (i.e. m = 4 hours, 3 sub-intervals for Type I/II error probability approximation). Therefore, as shown in Fig. 7.2, the optimally selected P_b value is obtained by maximizing $E[C|A]$. The maximum is denoted with a square, ($\square$), and the corresponding axes is to the right. The false alarm and missed detection probabilities are annotated as shown as a function of P_b, and the corresponding axis is to the left. The Type I/II error probabilities corresponding to the selection of P_b that result from choosing this optimal maximum $E[C|A]$ value are denoted with asterisks. Note that the alarm probability corresponding to the selection of P_b is very small. Its axes can also be read from the left. In fact, the alarm will rarely sound with such a small alarm probability, and even so the missed detection probability is nearly one. Therefore, as a practical measure, using the $E[C|A]$ design criterion does not prove to be useful.

In order for an optimal alarm system to have any utility, that is, have any clear economic value, there must be some way of showing how it can save money. This is possible is by considering both energy and complaint costs. Since energy costs have implications on optimal control, we will defer a discussion of the energy cost implications of optimal alarm systems until Sec. 7.3 is presented.

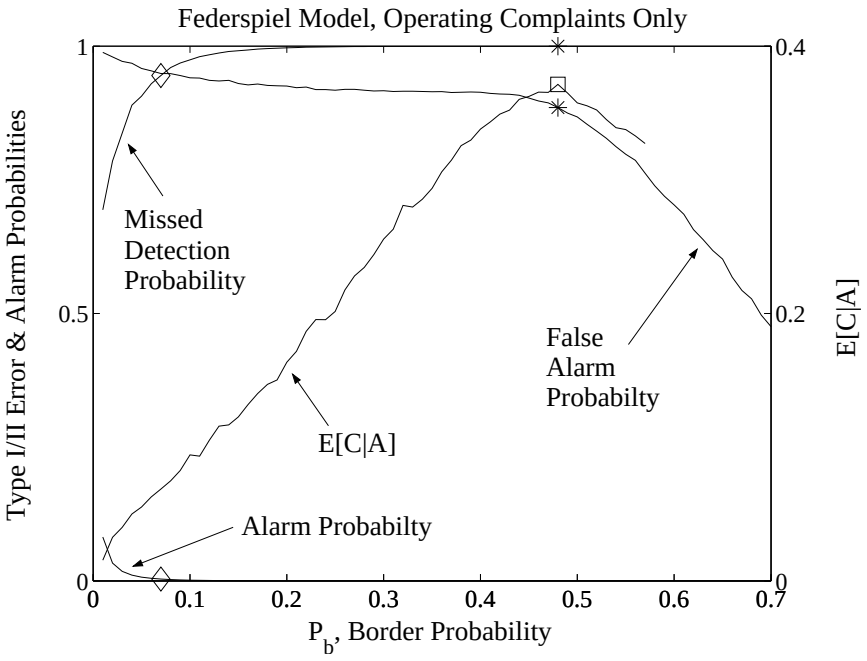

Figure 7.2: Selecting the Optimal P_b Value

However, we also may assume that any alarms triggered by the system may result in control actions ranging from no response to immediate response by facility managers, based on their knowledge of the energy budget. This leaves the energy cost implications of optimal alarm systems completely in the hands of those who set the building thermostat setting control policy. That still leaves the question of complaint costs open, and its bearing on the utility of the optimal alarm system alone, without any control action tied to it. We can categorize response to alarm events as being automated or unautomated. It is important to make this distinction because if the response is unautomated, the alarm system utility will be different than if it is automated.

If we consider the case in which the response is unautomated, it becomes clear that there is no inherent utility in the optimal alarm system under certain conditions. Let's assume that facility managers will *always* respond to complaint or alarm events by dispatching technicians, hence incurring labor costs *regardless* of the energy implications. As such, complaints must be responded

to whether an alarm sounds or not (i.e. whether the there was a missed or correct detection). In other words we "pay now or pay later" when it comes to the utility of an optimal alarm system. If there were no alarm system, or if there is a missed detection with an alarm system, we will eventually pay for the response to the complaint anyway. If there is a correct alarm, and we take action now instead of later when the complaint actually occurs, we are still paying the labor charges for the facility technician to respond !

However, in the case of a *false* alarm, any money spent on dispatching facility technicians is money wasted. Therefore, using an optimal alarm system under these conditions may actually have a negative utility. If we want to use this optimal alarm system, the only thing we are gaining is an advantage in the thermal comfort of building occupants, not utility in the monetary sense. Unless we attach a dollar value to discomfort (which has been investigated in the literature by Mozer et.al. [52, 53]), this alarm system will not increase utility. In a sense, if the optimal alarm system is completely decoupled from energy considerations (the optimal control problem), and acts alone in an unautomated fashion, our original assumption of minimizing false alarms was a good one. Doing so will mitigate the decrease in utility of the optimal alarm system.

If we consider the case in which the response is automated, the utility of the optimal alarm system changes. In fact, simply minimizing false alarms may not be the best strategy in this case. Automation of the response implies that when an alarm occurs, there may be some centralized operation point as there typically are in large commercial DDC systems, where an operator can simply "push a button" in order to initiate a "semi-automated" response to the alarm. Or the response can by fully automated, in which there is no operator action required at all. However, in either case, there is no need to dispatch a facility technician, and therefore no labor charges are incurred. Therefore, the utility of the alarm system must be tied to energy considerations, because the complaint cost as defined thus far has been eliminated via automation. Hence we defer further discussion of this case until the optimal control results have been presented. It may still be of use to quantify a metric that considers both complaint and energy costs as originally defined. This allows

us to attribute *some* utility to complaints avoided by an automated preemptive alarm-based control action, in order to quantify the less tangible thermal comfort advantage.

Because we will almost always need to consider energy in response to alarm events even in the unautomated case, practically, we should try to find a way to balance missed detections with false alarms more evenly in the selection of P_b. We know that designing the optimal alarm system strictly to minimize false alarms results in a lack of utility. As such, we shall investigate alternate potential cost function candidates. Although there are many, as discussed in Chap. 5, we'll focus on the best tradeoff between Type I/II error probabilities: the minimax design criteria. Even though it may result in *seemingly* very high false alarm and missed detection rates, the optimal P_b value can be determined very easily using this method. From Fig. 7.2 this can clearly be discerned. The diamond shaped annotations in the figure indicate the even tradeoff between missed detection and false alarm probabilities. Furthermore, the alarm probability, also annotated with a diamond, is not so low that alarms will be too infrequent for any practical purposes. Therefore, for the remainder of this section, we will use the minimax method as the primary optimal alarm design criterion. Interpretation of the high false alarm/missed detection rate will be discussed in greater depth later in the implementation portion of this section.

7.2.2 Practical Issues in the Design of Optimal Alarm Systems

In order to design the optimal alarm system, and to arrive at a unique solution for the optimal P_b value, there are some basic data requirements that need to be met. When considering the building temperature process model based purely on 2^{nd} order statistics, we can think of obtaining the optimal P_b as a function of those statistics in addition to the complaint level statistics. In fact, that is actually how it is found during the design stage. The function arguments, therefore would be the parameters in Table 2.3 and the building parameters in Table 2.4. The parameters in Table 2.3 can be found by formulae shown in Eqns. 2.6-2.9, and the remaining measurable statistics can be computed by methods as described in Federspiel [19].

Some of the other parameters required to design the optimal alarm system are as follows:

- The period of interest, i.e whether the null hypothesis is for operating complaints only, or both arrival and operating complaints.
- The length of the prediction window, whose default value is the finite horizon chosen in Chap. 6 of $m = 2$ hours.
- If the null hypothesis is for both arrival and operating complaints, the default look-ahead arrival period is $d = 1$-hour.
- Whether the design is for hot or cold events.
- The sampling interval of discretization, with a default value of $T_s = 20$ min.
- The number of points to use in all numerical integrations performed, default of $N = 360$.
- The design method of choice (i.e. Minimax vs. $E[C|A]$ method, etc.). We will default on the minimax choice for reasons as covered earlier in this section.
- All remaining relevant approximation parameters, according to the approximation method of choice (i.e. 2D, sub-interval, multi-dimensional, etc.) The default method will be the sub-interval approximation using 3 sub-intervals, and for "both arrival and operating complaints", the subintervals are modelled as downcrossings.

When considering design of the optimal alarm system for the first-principles-based model, the requirements change. Recall that the main reason for considering an optimal alarm system based on first principles was to allow for linking the optimal alarm system to optimal control. The statistics-based model can only be implemented on a stand-alone basis, solely for unautomated operation. The primary differences are that instead of the 2^{nd} order building temperature statistics, the modelling details need to be provided as arguments to the function. For example, the state-space

matrices, the input noise covariance matrix, the steady-state, and the measurement noise variance are the required modelling details for the design of the first-principle based optimal alarm system.

There are several ways to determine the actual optimal P_b value. The simplest way is to perform a grid search, and selecting a fine enough resolution such that $P_b^{opt} \in [0, 1]$ yields a unique solution. This is also heavily dependent on the design method of choice (i.e. minimax vs. $E[C|A]$ method, etc.) It is also possible, and simpler to use a well-known search method, i.e. a scalar bounded nonlinear function minimization routine that can be easily coded. The method used was an algorithm based on a golden section search and parabolic interpolation [23], for which there are implementations in more than one language. A key observation when designing any optimal alarm system is that the design procedure may fail to converge to a reasonable optimal P_b over the stated search range of P_b's. For our particular application, P_b^{opt} is often very close to the left end of the search range. In order to address this, we may simply reduce the search range of $P_b^{opt} \in [0, 1]$ to some smaller range, i.e. $P_b^{opt} \in [0, 0.5]$ or $P_b^{opt} \in [0, 0.2]$.

One final consideration has to do with the real-time status of the optimal alarm design. Because the design is *not* an on-line real-time implementation, there is no particular premium on computational efficiency. There has been a recent study on optimal alarm system implementation that is theoretically suitable for an adaptive, on-line implementation by Antunes et.al. [2]. This implementation uses a different paradigm of Bayesian prediction, where the parameters are learned during system execution, but it was found that there was much work to be done on improving computational efficiency. In our case, the computing time should be kept to a minimum, even though the design is performed off-line. Statistics should be "refreshed" on a periodic basis, and it should not take an unreasonable amount of "off-line time" to design an optimal alarm system, based upon these new statistics. The "refresh period" should be commensurate with significant changes in model behavior, but can completely be based upon heuristics. A good rule of thumb for re-design of an optimal alarm system is upon any major change of HVAC system equipment, new tenant occupation, or change of season.

7.2.3 Practical Issues in the Implementation of Optimal Alarm Systems

With coverage of the practical issues in the design stage complete, it is now necessary to further the discussion by addressing implementation issues. The design stage was focused on finding an optimal P_b value. Now, during the implementation stage we assume that the optimal P_b value has already been selected, and all that remains is to provide the tools for the optimal alarm system's execution. There are several different methods for implementing an optimal alarm system, based upon the approximations available to us. However, the primary purpose of the approximations was to reduce the computational burden during the design stage, in lieu of simulating and "counting" to obtain the required probabilities. As an added benefit, by making these approximations we provide mathematical tractability to finding a semi-infinite hyper-rectangular bound on the convex alarm region. Due to the grid search, or golden section search required during the design stage, computing the exact alarm region or even the approximations would prove to very computationally intensive at implementation time. This is further exacerbated by the exponential number of computations required to find the Type I/II error probabilities due to the inlcusion/exclusion rule.

During the implementation stage, these computational difficulties are mitigated by the fact that the alarm region itself does not need to be parameterized by the time-indexed Kalman prediction variables. Furthermore, no probabilities actually need to be computed, since they were most important during the design stage. Rather, we only need to know if the alarm region is entered into for the current time step. Therefore, only a single binary decision needs to be performed, rather than a search over an range of values having an exponential computational effort. As part of the binary decision that needs to be made, we have several different options available to us. The first option is to use the exact alarm condition itself as the binary test. An example would be the condition shown as Eqn. 3.2 from Chap. 3, for operating complaints only. This would require the numerical computation of the corresponding integral only. The computational burden of this would not at all limit the real-time performance of the system, even if the sampling interval were a quarter of the one used during simulations (i.e. $T_s = 5$ min). Most single computations of this sort take on

the order of hundredths of a second.

It is still possible to reduce the computational burden at implementation time even further by using one of the approximations available during the design stage. Just as in the design stage, the approximation chosen has a computational burden that scales with the method selected. However, none of the methods require any numerical integration. Rather, since the approximations were all found with rectangular or hyper-rectangular semi-infinite regions, the alarm will sound any time the regions are entered, as defined by simple binary inequality checks. As a rule of thumb, if using these approximations during the implementation stage, it is best to apply the same approximation that was used during design. Moving to a more accurate method is always acceptable, however, so the most accurate binary check would be to use the exact alarm condition.

In addition to real-time considerations of the optimal alarm system, it is also of interest to consider the application itself. Specifically, we are interested in when the optimal alarm systems should be activated, and/or deactivated, based upon a typical schedule and critical periods throughout the course of a normal working day in a commercial building. A sample schedule is provided in Table 7.1.

HVAC System start	4:30 am
Activate Arrival (& Operating) Alarm Systems	5:30 am
Start of Day/Deactivate Arrival (& Operating) Alarm Systems	7:30 am
Start of First Arrival (& Operating) Complaint Period	6:30 am
End of First Arrival (& Operating) Complaint Period	8:30 am
Start of Last Arrival (& Operating) Complaint Period	8:30 am
End of Last Arrival (& Operating) Complaint Period	10:30 am
Activate Operating Only Alarm Systems	8:30 am
Deactivate Operating Only Alarm Systems	3:00 pm
Start of First Operating Only Complaint Period	8:30 am
End of First Operating Only Complaint Period	10:30 am
Start of Last Operating Only Complaint Period	3:00 pm
End of Last Operating Only Complaint Period	5:00 pm
End of Day/HVAC System shutdown	5:00 pm

Table 7.1: Schedule of Critical Periods And Events in Prototype Building

Shown in Table 7.1 are all of the standard events that occur throughout the day. It begins

with the HVAC system start-up, which typically happens early in the morning (4:30 am) well before the majority of the building occupants arrive, giving the zone time to reach a reasonable level of thermal comfort. The activation of the arrival and operating alarm systems (both hot and cold) should take place at 5:30 am. The reason behind this is that the look-ahead arrival period is $d = 1$-hour, and the finite horizon prediction window was chosen in to be $m = 2$ hours. Therefore, the "first" arrival and operating complaint period spans from 6:30 am to 8:30 am. This alarm system is continuously in operation until deactivated, and the "operating only" alarm system takes over. So the activation period continues to operate until 7:30 am to monitor for arrival complaints during the "last" arrival and operating complaint period spanning from 8:30 am to 10:30 am.

At 8:30 am, the operating only alarm system takes over, continuing to check for operating complaints for the same period that the previous alarm system monitored for, spanning from 8:30 am to 10:30 am. Notice that there is no look-ahead arrival period (i.e. $d = 0$). In this way, there is no loss in continuity of monitoring for complaint events, even though there is a 1-hour period from 7:30 am to 8:30 am where there is no active alarm system in operation. The "operating only" alarm system will continue in operation until it is deactivated at 3:00 pm, at which time it monitors for operating complaints spanning from 3:00 pm to 5:00 pm. 5:00 pm also denotes the time for the end of a typical workday. It is also at this time that the HVAC system is shutdown until the next morning at 4:30 am.

It is important to note that the optimal alarm systems operating during the time periods rely heavily on the results of Kalman filtering and prediction. Recall that the Kalman filter runs continuously throughout the 24-hour period, so that there is no "burn in" period necessary for convergence, other than on initial set-up. If the Kalman filter were scheduled to start and stop at the times the respective alarm systems were activated and deactivated, there would not be enough time for the Kalman filter to reach steady state. Furthermore, there is a second "burn-in" phenomenon that relies on the assumptions made concerning the stationarity and Gaussianity of the Kalman filter. This was mentioned briefly before in Chap. 3.

Recall the "counting" method, where the statistics of Type I/II error probabilities are obtained by using ergodicity. Essentially, simulations may be performed to obtain an estimate of these probabilities. They are obtained by running a Kalman predictor, and counting the number of correct/false alarms and correct/missed detections, for a long enough time until their relative frequencies converge to limiting probability values. Hence, we can characterize this as a "burn-in" phenomenon. However, the modelling assumptions made concerning the Kalman filter must be close to reality. Otherwise, the convergence will with take longer than normal, or convergence to values that are different than the theoretically computed ones. Theoretical computation in this sense refers to the numerical computation of the Type I/II error probabilities. When a model used as the basis for Kalman filtering and prediction is used obtain these probability via simulation, we should have "proper convergence." Proper convergence can be defined as convergence to the Type I/II error probabilities that were computed via numerical integration, using a practical approximation during the design stage corresponding to the optimal P_b value. The actual simulated model output, and *real* data from a real system will inevitably produce different results than a model used as the basis for Kalman filtering and prediction. As a result, we must investigate the sub-optimality of these results based on the convergence properties of using real and simulated data.

Recall from the Chap. 6 that the model-based simulation simply marches forward according to the available TMY weather data and simulated internal heat loads using Eqn. 6.5. However, when performing Kalman filtering and prediction for the optimal alarm system, Gaussian assumptions about u_k and $\mathbf{w}_k$ were made. The Gaussian assumption on u_k accounts for the model's non-stationary night and weekend periods by assuming that $u_k \sim \mathcal{N}(0, \sigma^2_{T_{out}})$. The Gaussian assumption on $\mathbf{w}_k$ accounts for the mean and variance of the non-stationary and non-Gaussian input disturbance using the Gaussian paradigm, so that $\mathbf{w}_k \sim \mathcal{N}(\bar{\mathbf{w}}_k, \mathbf{W})$. Now in addition to the modelling differences there are other factors that may influence the convergence properties of the Type I/II error probabilities. One factor addresses the issue of the alarm systems' operational schedules. They are not continuously operating, but only at the periods listed in Table 7.1. Therefore, time to

convergence is periodically "interrupted." Another factor addresses alarm system implementation issues. We may be using a different implementation than was used at the design stage. For example, we may have designed the optimal alarm system using the sub-interval approximation method, but during the implementation stage we used the exact alarm condition. This will also lend itself to deviation from proper convergence.

So far, we've focused completely on the first-principles based model. However, if we are considering the optimal alarm system alone, without any subsequent automated control action, the same investigation of sub-optimality needs to take place. In this case, Kalman filtering and prediction is based on the simple 2^{nd}-order statistics-based model. The resulting Type I/II error probability convergence will be studied by considering the different circumstances described thus far. Those circumstances are summarized as follows:

- Design mismatch vs. no mismatch with the implementation method.
- Real vs. simulated data.
- With and without scheduling. "Without scheduling" implies that all optimal alarm systems (hot/cold, both arrival & operting/operating only) will operate continuously for all 24 hours of the day. "With scheduling" implies that the optimal alarm systems will operate according to the schedule shown in Table 7.1.

Tables 7.2-7.29 illustrate the convergence properties, including proper convergence or the lack of it for all of the candidate buildings described in Sec. 7.1. In these tables the optimal alarm *design* was based on the simple 2^{nd}-order statistics-based model. The design parameters were the default ones described in the list earlier in this section. Below, Tables 7.2- 7.5 represent the results for Building A when using the statistics-based model for simulation. The results also represent continuous 24-hour operation of the alarm systems.

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	71	71
Simulated False Alarm Probability	88.73	88.73
Numerical False Alarm Probability	96.06	96.06
Simulated Missed Detection Count	109	41
Simulated Missed Detection Probability	92.66	80.49
Numerical Missed Detection Probability	95.7	95.7
Optimal P_b	0.0214	0.0214
Run Time (years)	6.09	3.25

Table 7.2: Table of Statistical Convergence for Building A, Hot Arrival & Operating Alarm System, Statistics-Based Model

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	1123	2071
Simulated False Alarm Probability	94.3	96.43
Numerical False Alarm Probability	92.36	92.36
Simulated Missed Detection Count	1281	659
Simulated Missed Detection Probability	95	88.77
Numerical Missed Detection Probability	92.18	92.18
Optimal P_b	0.0333	0.0333
Run Time (years)	6.09	3.25

Table 7.3: Table of Statistical Convergence for Building A, Cold Arrival & Operating Alarm System, Statistics-Based Model

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	6	40
Simulated False Alarm Probability	50	90
Numerical False Alarm Probability	99.18	99.18
Simulated Missed Detection Count	48	23
Simulated Missed Detection Probability	93.75	82.61
Numerical Missed Detection Probability	97.65	97.65
Optimal P_b	0.026	0.026
Run Time (years)	6.09	3.25

Table 7.4: Table of Statistical Convergence for Building A, Hot Operating Only Alarm System, Statistics-Based Model

Looking at the results presented in Tables 7.2- 7.5, we see that there are 4 different tables to represent the 4 different alarm systems needed during system operation. There are two columns for each table, representing the two different implementation types to be compared. In this case, the default approximation (3 sub-intervals, downcrossings for both arrival & operating complaints) is being compared to the exact alarm region. The run time for both methods, respectively, is 6.09 and 3.25 years. Both should be sufficient time for convergence to the limiting false alarm and

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	246	1158
Simulated False Alarm Probability	96.34	97.75
Numerical False Alarm Probability	94.61	94.61
Simulated Missed Detection Count	742	388
Simulated Missed Detection Probability	98.79	93.3
Numerical Missed Detection Probability	94.422	94.422
Optimal P_b	0.0424	0.0424
Run Time (years)	6.09	3.25

Table 7.5: Table of Statistical Convergence for Building A, Cold Operating Only Alarm System, Statistics-Based Model

missed detection probabilities. The numerically computed probability values are shown, and can be compared directly to the simulated probability values. For the simulated runs, we show the number of counts recorded as well, to provide a gauge as to the accuracy of the simulation-based probabilities. The errors themselves are not included in the table (they can be computed accurately for simulated results and also numerical results to a lesser extent). This is to allow for a more clear presentation of the key information.

For the first column, when using the default approximation, we should expect that the simulated probability valued converge to the numerically computed ones. The reason for this is that the design was based on the same default approximation and statistics-based model that was used during simulation. Also, the results represent continuous 24-hour operation of all alarm systems. Demonstration of proper convergence within "ϵ," however, is not likely even under these theoretically optimal conditions. This is mainly due to the fact that the run time would have to continue *ad infinitum* until the counts for missed detections and false alarms grew infinitely large. We see that in Table 7.3, where the counts are largest, the false alarm probabilities are respectively 94.3 and 92.36 for the simulated and numerically computed results. Similarly, the missed detection probabilities are respectively 95 and 92.18 for the simulated and numerically computed results. This is a fairly close agreement, and looking at the remaining Tables 7.2- 7.5, where the counts are less, we see that the agreement gets worse and worse as the counts decrease. Therefore, if we ran the simulations for longer, the agreement between simulated and numerical computed probabilities would improve

considerably. Hence the simulated probability results would approach the numerically computed results within a tolerance of the error involved in each method.

Now, we'd like to determine how much the results change when we use a different implementation method. By moving to the second column in Tables 7.2- 7.5 above, we see the exact alarm region is used in lieu of the default approximation. Therefore, the limiting probabilities achieved by running simulations should show some sub-optimality and exhibit less than perfect convergence. Again, we see that in Table 7.3, where the counts are largest, the false alarm probabilities are respectively 96.43 and 92.36 for the simulated and numerically computed results. Similarly, the missed detection probabilities are respectively 88.77 and 92.18 for the simulated and numerically computed results. Therefore, the simulation-based probabilities have not changed considerably when using the new implementation method. The differences are so small that is it is unclear whether they are due to the combination of simple simulation and numerical integration error, or due to the anticipated lack of proper convergence from sub-optimality.

As such, we will now turn to more significant changes that may have a greater contribution to sub-optimality. Specifically, when we use the real building data used to design the optimal alarm system, we will have a better measure of its practical usefulness. Below, Tables 7.6- 7.9 represent the results for Building A when using real data. The results also represent optimal alarm system implementations which are based upon the exact alarm region.

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	83	135
Simulated False Alarm Probability	97.59	98.52
Numerical False Alarm Probability	96.06	96.06
Simulated Missed Detection Count	33	8
Simulated Missed Detection Probability	93.94	75
Numerical Missed Detection Probability	95.7	95.7
Optimal P_b	0.0214	0.0214
Run Time (years)	3.63	9.4

Table 7.6: Table of Statistical Convergence for Building A, Hot Arrival & Operating Alarm System, Real Data

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	1591	554
Simulated False Alarm Probability	94.34	99.1
Numerical False Alarm Probability	92.36	92.36
Simulated Missed Detection Count	655	50
Simulated Missed Detection Probability	86.26	50
Numerical Missed Detection Probability	92.18	92.18
Optimal P_b	0.0333	0.0333
Run Time (years)	3.63	9.4

Table 7.7: Table of Statistical Convergence for Building A, Cold Arrival & Operating Alarm System, Real Data

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	77	28
Simulated False Alarm Probability	97.4	100
Numerical False Alarm Probability	99.18	99.18
Simulated Missed Detection Count	19	15
Simulated Missed Detection Probability	89.47	100
Numerical Missed Detection Probability	97.65	97.65
Optimal P_b	0.026	0.026
Run Time (years)	3.63	9.4

Table 7.8: Table of Statistical Convergence for Building A, Hot Operating Only Alarm System, Real Data

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	698	57
Simulated False Alarm Probability	97.99	100
Numerical False Alarm Probability	94.61	94.61
Simulated Missed Detection Count	384	29
Simulated Missed Detection Probability	96.35	100
Numerical Missed Detection Probability	94.42	94.42
Optimal P_b	0.0424	0.0424
Run Time (years)	3.63	9.4

Table 7.9: Table of Statistical Convergence for Building A, Cold Operating Only Alarm System, Real Data

As before, looking at the results presented in Tables 7.6- 7.9, we see that there are 4 different tables to represent the 4 different alarm systems needed during system operation. There are two columns for each table, representing the two different scheduling types to be compared. In this case, continuous 24-hour operation of the alarm systems is being compared to the schedule provided in Table 7.1. The run time for both methods, respectively, is 3.63 and 9.4 years. Both should be sufficient time for convergence to the limiting false alarm and missed detection probabilities.

However, the run-times shown are based on repeated trials of the same dataset since there is a limited 125.82 day period available for Building A.

Recall that Tables 7.6- 7.9 represent more significant changes that may have a greater contribution to sub-optimality. Therefore, we should expect a reasonable deviation from proper convergence. Let's use Table 7.7 as a guide, where the counts are largest again for the cold arrival & operating alarm system. When considering the first column, for 24 hour continuous operation, the false alarm probabilities are respectively 94.34 and 92.36 for the simulated and numerically computed results. Similarly, the missed detection probabilities are respectively 86.26 and 92.18 for the simulated and numerically computed results. For the second column, when considering scheduling of the alarm system according to Table 7.1, the false alarm probabilities are respectively 99.1 and 92.36 for the simulated and numerically computed results. Similarly, the missed detection probabilities are respectively 50 and 92.18 for the simulated and numerically computed results.

We see that the results do not change significantly when using real data, for continuous alarm system operation. However, when applying the realistic schedule, they *do* change significantly. Unfortunately, this is more an artifact of the small number of counts (especially for missed detections at only 50) rather than from a deviation from perfect convergence due to sub-optimality. As such when we investigate the remaining two buildings for this particular set of circumstances, we will increase the run-time considerably, even though it is already over 9 years in this case.

The same findings for all the other cases which have not been influenced by a small missed detection or false alarm count can be verified. This can be done by performing the same investigation for the remaining two candidate buildings, Buildings C & F. Therefore, Tables 7.10- 7.13 represent the results for Building F when using the statistics-based model for simulation. These results also represent continuous 24-hour operation of the alarm systems.

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	234	110
Simulated False Alarm Probability	86.75	100
Numerical False Alarm Probability	95.4	95.4
Simulated Missed Detection Count	186	135
Simulated Missed Detection Probability	83.33	100
Numerical Missed Detection Probability	93.91	93.91
Optimal P_b	0.0242	0.0242
Run Time (years)	8.18	5.27

Table 7.10: Table of Statistical Convergence for Building F, Hot Arrival & Operating Alarm System, Statistics-Based Model

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	1459	3831
Simulated False Alarm Probability	91.84	95.09
Numerical False Alarm Probability	89.82	89.82
Simulated Missed Detection Count	1222	702
Simulated Missed Detection Probability	90.26	73.22
Numerical Missed Detection Probability	87.23	87.23
Optimal P_b	0.0344	0.0344
Run Time (years)	8.18	5.27

Table 7.11: Table of Statistical Convergence for Building F, Cold Arrival & Operating Alarm System, Statistics-Based Model

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	120	37
Simulated False Alarm Probability	95	100
Numerical False Alarm Probability	96.82	96.82
Simulated Missed Detection Count	49	59
Simulated Missed Detection Probability	87.76	100
Numerical Missed Detection Probability	96.18	96.18
Optimal P_b	0.0336	0.0336
Run Time (years)	8.18	5.27

Table 7.12: Table of Statistical Convergence for Building F, Hot Operating Only Alarm System, Statistics-Based Model

Observing the results for Building F, when using the statistics-based model for simulation, and representing continuous 24-hour operation, slightly different conclusions can be made compared to Building A. They are listed as follows:

- For the implementation method that uses the default approximation, if we ran the simulations for longer, the agreement between simulated an numerical computed probabilities would improve considerably. Hence the simulated probability results would approach the numeri-

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	966	2897
Simulated False Alarm Probability	95.34	96.96
Numerical False Alarm Probability	95.74	95.74
Simulated Missed Detection Count	661	404
Simulated Missed Detection Probability	93.19	78.22
Numerical Missed Detection Probability	93.37	93.37
Optimal P_b	0.0396	0.0396
Run Time (years)	8.18	5.27

Table 7.13: Table of Statistical Convergence for Building F, Cold Operating Only Alarm System, Statistics-Based Model

cally computed results within a tolerance of the error involved in each method. In this case, with Building F, we see that there is a particularly good agreement in Table 7.13 for the cold operating only alarm system, even for a 661 missed detection count during an 8 year run.

- For the implementation method that uses the exact alarm region, the differences for this case are a bit larger in Building F than with Building A. For most cases shown in Tables 7.10-7.13, it is clear that the differences are not due to the combination of simple simulation and numerical integration error. The only exception might be in Table 7.13, where the false alarm probabilities are very similar. The differences in the remaining cases are most likely due to the anticipated lack of proper convergence from sub-optimality, or not enough false alarm/missed detection counts as in Tables 7.10 and 7.12.

Now Tables 7.14- 7.17 represent the results for Building F when using real data. These results also represent optimal alarm system implementations which are based upon the exact alarm region.

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	0	0
Simulated False Alarm Probability	N/A	N/A
Numerical False Alarm Probability	95.4	95.4
Simulated Missed Detection Count	56	8
Simulated Missed Detection Probability	100	100
Numerical Missed Detection Probability	93.91	93.91
Optimal P_b	0.0242	0.0242
Run Time (years)	5.51	9.54

Table 7.14: Table of Statistical Convergence for Building F, Hot Arrival & Operating Alarm System, Real Data

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	8458	693
Simulated False Alarm Probability	92.55	94.95
Numerical False Alarm Probability	89.82	89.82
Simulated Missed Detection Count	1103	62
Simulated Missed Detection Probability	42.88	43.55
Numerical Missed Detection Probability	87.23	87.23
Optimal P_b	0.0344	0.0344
Run Time (years)	5.51	9.54

Table 7.15: Table of Statistical Convergence for Building F, Cold Arrival & Operating Alarm System, Real Data

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	0	0
Simulated False Alarm Probability	N/A	N/A
Numerical False Alarm Probability	96.82	96.82
Simulated Missed Detection Count	21	0
Simulated Missed Detection Probability	100	N/A
Numerical Missed Detection Probability	96.18	96.18
Optimal P_b	0.0336	0.0336
Run Time (years)	5.51	9.54

Table 7.16: Table of Statistical Convergence for Building F, Hot Operating Only Alarm System, Real Data

The run-times shown are now based on repeated trials of the same dataset with the limited 45.16 day period available for Building F. As evidenced in Tables 7.14- 7.17, there are some large differences in the results for 24-hour operation of the alarm systems. For example, in Tables 7.14 and 7.16, for all hot alarm events, there are so few missed detections and false alarms that the simulated probabilities are sometimes incalculable. When they are, it is often at the extreme end of 100% for missed detections. This is largely due to the short repeated time span (45.16 days) and

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	8017	1463
Simulated False Alarm Probability	96.26	97.33
Numerical False Alarm Probability	95.74	95.74
Simulated Missed Detection Count	581	181
Simulated Missed Detection Probability	48.36	78.45
Numerical Missed Detection Probability	93.37	93.37
Optimal P_b	0.0396	0.0396
Run Time (years)	5.51	9.54

Table 7.17: Table of Statistical Convergence for Building F, Cold Operating Only Alarm System, Real Data

therefore limited range of the real data, which is insufficient to generate enough actual complaint events. Other major deviations include the missed detection probabilities for Tables 7.15 and 7.17, for all cold alarm systems. Although these same large deviations are not evident in the respective false alarm probabilities, they stem from the fact that the real data is not well characterized by its aggregate statistical representation.

Furthermore, there is no significant improvement in the small number of counts when using the real schedule in the second column. Much of same behavior is present as in continuous operation, in consideration of the large deviations, extremes, and incalculability of the probabilities. Therefore, it is still hard to discern whether the differences between simulated and numerically computed probabilities is due to lack of sufficient run-time rather than from a deviation from perfect convergence due to sub-optimality. Unfortunately, when investigating Building C, we may run into the same issues because the run-time associated with the sample data set was only 82.2 days. As a result, instead of relying on real data to investigate deviations from perfect convergence due to sub-optimality, we will use a model that is very qualitatively similar to the real data. In this way, we can extend the range of the values in the dataset, and use as long a run-time is necessary without having to worry about re-using an old dataset.

We've left Building C for last because it is the candidate that was used as the prototype building/zone for comparison to a first-principles model as discussed in Sec. 7.1. Therefore, we will now perform the same analysis as was done for Buildings A and F, and follow it up with a similar

examination when using the first-principles based model. The optimal alarm design will still be based on the simple 2^{nd}-order statistics-based model. However, we will follow that up by a study of optimal alarm *design* that is based on the first-principles based model.

Tables 7.18- 7.21 represent the results for Building C when using the statistics-based model for simulation. These results also represent continuous 24-hour operation of the alarm systems.

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	101	27
Simulated False Alarm Probability	87.13	81.48
Numerical False Alarm Probability	92.45	92.45
Simulated Missed Detection Count	148	58
Simulated Missed Detection Probability	91.22	91.38
Numerical Missed Detection Probability	93.67	93.67
Optimal P_b	0.0359	0.0359
Run Time (years)	11.93	8.39

Table 7.18: Table of Statistical Convergence for Building C, Hot Arrival & Operating Alarm System, Statistics-Based Model

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	4183	9439
Simulated False Alarm Probability	84	90.21
Numerical False Alarm Probability	82.87	82.87
Simulated Missed Detection Count	4621	3410
Simulated Missed Detection Probability	85.52	72.9
Numerical Missed Detection Probability	84.77	84.77
Optimal P_b	0.0697	0.0697
Run Time (years)	11.93	8.39

Table 7.19: Table of Statistical Convergence for Building C, Cold Arrival & Operating Alarm System, Statistics-Based Model

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	85	46
Simulated False Alarm Probability	98.82	95.65
Numerical False Alarm Probability	98.79	98.79
Simulated Missed Detection Count	67	18
Simulated Missed Detection Probability	98.51	88.89
Numerical Missed Detection Probability	98.42	98.42
Optimal P_b	0.0326	0.0326
Run Time (years)	11.93	8.39

Table 7.20: Table of Statistical Convergence for Building C, Hot Operating Only Alarm System, Statistics-Based Model

Implementation Type	Default Approximation	Exact Alarm Region
Simulated False Alarm Count	2926	8455
Simulated False Alarm Probability	91.18	95.16
Numerical False Alarm Probability	92.5	92.5
Simulated Missed Detection Count	2482	1857
Simulated Missed Detection Probability	89.61	77.98
Numerical Missed Detection Probability	88.57	88.57
Optimal P_b	0.0742	0.0742
Run Time (years)	11.93	8.39

Table 7.21: Table of Statistical Convergence for Building C, Cold Operating Only Alarm System, Statistics-Based Model

Observing the results for Building C, when using the statistics-based model for simulation, and representing continuous 24-hour operation, unique conclusions can be made, and are listed as follows:

- In this case, with Building C, we see that there is a particularly good agreement between the simulated and numerically computed probabilities for all alarm systems. Therefore, the run times that have been used are sufficient.

- For the implementation method that uses the exact alarm region, the differences for this case (Building C) are more pronounced than in Building A, but about the same as in Building F. In most cases shown in Tables 7.18- 7.21, it is clear that the differences are not due to the combination of simple simulation and numerical integration error. This is a result of using a long enough simulation run-time. Therefore, the differences between the simulated and numerically computed probabilities are most likely due to the anticipated lack of proper convergence from sub-optimality.

Tables 7.22- 7.25 represent the results for Building C when using real data. These results also represent optimal alarm system implementations which are based upon the exact alarm region.

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	40	109
Simulated False Alarm Probability	100	100
Numerical False Alarm Probability	92.45	92.45
Simulated Missed Detection Count	40	0
Simulated Missed Detection Probability	100	N/A
Numerical Missed Detection Probability	93.67	93.67
Optimal P_b	0.0359	0.0359
Run Time (years)	9.04	24.66

Table 7.22: Table of Statistical Convergence for Building C, Hot Arrival & Operating Alarm System, Real Data

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	8831	1390
Simulated False Alarm Probability	86.67	99.14
Numerical False Alarm Probability	82.87	82.87
Simulated Missed Detection Count	3982	136
Simulated Missed Detection Probability	70.44	91.18
Numerical Missed Detection Probability	84.77	84.77
Optimal P_b	0.0697	0.0697
Run Time (years)	9.04	24.66

Table 7.23: Table of Statistical Convergence for Building C, Cold Arrival & Operating Alarm System, Real Data

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	41	0
Simulated False Alarm Probability	100	N/A
Numerical False Alarm Probability	98.79	98.79
Simulated Missed Detection Count	19	28
Simulated Missed Detection Probability	100	100
Numerical Missed Detection Probability	98.42	98.42
Optimal P_b	0.0326	0.0326
Run Time (years)	9.04	24.66

Table 7.24: Table of Statistical Convergence for Building C, Hot Operating Only Alarm System, Real Data

Observing Tables 7.22- 7.25, it is evident that for continuous operation when using real data, looping the same dataset over and over even for a long time (over 9 years) with a repeat length of 82.2 days will not result in perfect convergence. This is due to the fact that there is a lack of sufficient frequency content in the data to bridge the discrepancy between the aggregate statistics used to design the alarm systems and the non-stationarity of the actual data. It is hard to make a judgement on the realistic schedule results shown in Tables 7.22- 7.25, due to the lack of sufficient

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	7914	0
Simulated False Alarm Probability	93.69	N/A
Numerical False Alarm Probability	92.5	92.5
Simulated Missed Detection Count	2008	44
Simulated Missed Detection Probability	75.15	100
Numerical Missed Detection Probability	88.57	88.57
Optimal P_b	0.0742	0.0742
Run Time (years)	9.04	24.66

Table 7.25: Table of Statistical Convergence for Building C, Cold Operating Only Alarm System, Real Data

false alarm and missed detection counts, with the notable exception of Table 7.23. This is true even though the total run length was close to 25 years !

Therefore, to address this limitation, we will now present the final results for Building C where the optimal alarm design is based on the simple 2^{nd}-order statistics-based model. However, the primary difference is that the actual implementation, or simulation will not be based on the same simple 2^{nd}-order statistics-based model used in design. In fact, the first-principles model discussed in Sec. 7.1 that is very qualitatively similar to the real data will be used in lieu of the real data. Consequently, there is no limitation on the span or range (of temperatures) in the data available, as was the case with real datasets. As such, Tables 7.26- 7.29 represent the results for Building C when the optimal alarm system implementations are based upon the exact alarm region.

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	722	7862
Simulated False Alarm Probability	100	99.95
Numerical False Alarm Probability	92.45	92.45
Simulated Missed Detection Count	18	30
Simulated Missed Detection Probability	100	86.67
Numerical Missed Detection Probability	93.67	93.67
Optimal P_b	0.0359	0.0359
Run Time (years)	3.86	43.27

Table 7.26: Table of Statistical Convergence for Building C, Hot Arrival & Operating Alarm System, Modelled Data

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	19184	12719
Simulated False Alarm Probability	80.69	99.49
Numerical False Alarm Probability	82.87	82.87
Simulated Missed Detection Count	4676	271
Simulated Missed Detection Probability	20.77	76.01
Numerical Missed Detection Probability	84.77	84.77
Optimal P_b	0.0697	0.0697
Run Time (years)	3.86	43.27

Table 7.27: Table of Statistical Convergence for Building C, Cold Arrival & Operating Alarm System, Modelled Data

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	772	0
Simulated False Alarm Probability	99.87	N/A
Numerical False Alarm Probability	98.79	98.79
Simulated Missed Detection Count	12	32
Simulated Missed Detection Probability	91.67	100
Numerical Missed Detection Probability	98.42	98.42
Optimal P_b	0.0326	0.0326
Run Time (years)	3.86	43.27

Table 7.28: Table of Statistical Convergence for Building C, Hot Operating Only Alarm System, Modelled Data

Schedule Type	24-hour	Table 7.1
Simulated False Alarm Count	18369	35
Simulated False Alarm Probability	91.56	94.29
Numerical False Alarm Probability	92.5	92.5
Simulated Missed Detection Count	2108	155
Simulated Missed Detection Probability	26.42	98.71
Numerical Missed Detection Probability	88.57	88.57
Optimal P_b	0.0742	0.0742
Run Time (years)	3.86	43.27

Table 7.29: Table of Statistical Convergence for Building C, Cold Operating Only Alarm System, Modelled Data

Similar to the previous cases when using real data, there are two columns for each of Tables 7.26- 7.29, representing the two different scheduling types to be compared. Continuous 24-hour operation of the alarm systems in the first column is being compared to scheduled operation in the next column. For continuous operation of the alarm system, it is clear that there are some examples where the simulated counts were not high enough to reduce the error in the corresponding false alarm or missed detection probability. Examples of this can be found in Table 7.26, and the

missed detection probability computation in Table 7.28, although the deviation from the numerically computed value is not unreasonable. However, the simulated probability counts in the remaining tables (for the cold alarm systems) are sufficient to reduce the error involved in the probabilities. The false alarm probabilities match fairly well for these remaining tables, however, the missed detection rates show a very strange anomaly. It appears that the simulated missed detections probabilities are far from their numerically computed counterparts, even though the missed detection counts are sufficiently high to reduce simulation error. This is due to the inherent non-stationarity of the data simulated by the model. Recall Fig. 7.1, in which the non-stationarity is quite evident. If more of the simulation were shown, the non-stationary night and weekend periods which are obvious in Fig. 7.1 would dip down lower and lower, and then eventually back up again as the seasons change and the outside temperature influence these "float" periods.

The influence of these non-stationary periods is clearly only present on the "cold" side, because during the normal occupied periods, the temperature is regulated by a controller. Here, the temperature is more "stationary", and therefore the "hot" side will not be affected by this anomaly. Therefore, the asymmetry in the agreements of simulated and numerically computed probability results between hot and cold can easily be attributed to this non-stationary phenomenon. The discrepancy between false alarm and missed detection probability agreement on the "cold" side can be explained by the non-stationarity as well. Actual cold complaints are logged during the unoccupied period due to continuous 24-hour operation of the alarm systems. Therefore, cold complaints will not be missed as often as they should due to the extreme downward excursions that take place on HVAC system shutdown. This explains the greatly reduced missed detection rate.

Now if we adhere to the schedule give in Table 7.1, this phenomenon is eliminated, and the missed detection rates should change significantly, as seen the second column of Tables 7.27 and 7.29. The remaining results are fairly similar to the results from the first column. However, there is a bit more deviation due to either the lack of events, a greater number of events, or due to the sub-optimality experienced by adhering to a schedule. For Tables 7.27 and 7.29, this schedule

operates only during stationary "operating only" periods that have a different statistical basis than the complete modelling dataset.

Before closing out this section and moving on to a practical discussion of optimal control in the next section, we provide the final results on optimal alarm that serve as its precursor. The optimal alarm *design* for all of the results presented thus far have been based on the simple 2^{nd}-order statistics-based model. This has been a useful investigation for optimal alarm systems that are not to be linked with subsequent automatic control responses. In fact, the overhead is quite less, and no modelling is necessary, only empirical data is required. However, in order to link an automated optimal control response, a first-principles model is necessary to design the optimal alarm system as well as the control algorithm. As such, we'd like to present the optimal alarm results of a system whose *design* was based on a first principles model.

Using the same framework as presented all along in this section, Tables 7.30-7.33 represent the statistics of optimal alarm systems whose designs were based on a first-principles model. Recall that certain assumptions were made in Sec. 6.2.3 of Chap. 6 in order to complete the design. These assumptions (mainly on Gaussianity and stationarity) were different than the actual model's real-time implementation. Rather, they are only used during optimal alarm system design. These discrepancies will cause sub-optimal results and therefore deviations from perfect convergence. As such, they will be analyzed as well, and in the table a new row, "Model Type" will be added, denoted by either "G-S Approx" for "Gaussian-Stationary Approximation, or "Real Model," for the actual first-principles model with the non-stationary and non-Gaussian input disturbances.

Model Type	G-S Approx	Real Model	Real Model
Implementation Type	Default Approximation	Default Approximation	Exact Alarm Region
Schedule Type	24-hour	24-hour	Table 7.1
Simulated Type I Count	233	0	36
Simulated Type I Probability	100	N/A	100
Numerical Type I Probability	95.84	95.84	95.84
Simulated Type II Count	153	14	2
Simulated Type II Probability	100	100	100
Numerical Type II Probability	95.71	95.71	95.71
Optimal P_b	0.00426	0.00426	0.00426
Run Time (years)	6.85	6.85	6.85

Table 7.30: Table of Statistical Convergence for Building C, Hot Arrival & Operating Alarm System, 1^{st}-principles

Model Type	G-S Approx	Real Model	Real Model
Implementation Type	Default Approximation	Default Approximation	Exact Alarm Region
Schedule Type	24-hour	24-hour	Table 7.1
Simulated Type I Count	4287	35909	3436
Simulated Type I Probability	99.93	80.02	99.45
Numerical Type I Probability	98.25	98.25	98.25
Simulated Type II Count	38	8836	30
Simulated Type II Probability	92.11	18.81	36.67
Numerical Type II Probability	98.09	98.09	98.09
Optimal P_b	0.0279	0.0279	0.0279
Run Time (years)	6.85	6.85	6.85

Table 7.31: Table of Statistical Convergence for Building C, Cold Arrival & Operating Alarm System, 1^{st}-principles

Model Type	G-S Approx	Real Model	Real Model
Implementation Type	Default Approximation	Default Approximation	Exact Alarm Region
Schedule Type	24-hour	24-hour	Table 7.1
Simulated Type I Count	410	0	18
Simulated Type I Probability	100	N/A	100
Numerical Type I Probability	97.92	97.92	97.92
Simulated Type II Count	60	6	16
Simulated Type II Probability	100	100	100
Numerical Type II Probability	96.59	96.59	96.59
Optimal P_b	0.00536	0.00536	0.00536
Run Time (years)	6.85	6.85	6.85

Table 7.32: Table of Statistical Convergence for Building C, Hot Operating Only Alarm System, 1^{st}-principles

Model Type	G-S Approx	Real Model	Real Model
Implementation Type	Default Approximation	Default Approximation	Exact Alarm Region
Schedule Type	24-hour	24-hour	Table 7.1
Simulated Type I Count	3669	0	96
Simulated Type I Probability	100	N/A	93.75
Numerical Type I Probability	98.25	98.25	98.25
Simulated Type II Count	24	3869	48
Simulated Type II Probability	100	100	87.5
Numerical Type II Probability	96.83	96.83	96.83
Optimal P_b	0.0363	0.0363	0.0363
Run Time (years)	6.85	6.85	6.85

Table 7.33: Table of Statistical Convergence for Building C, Cold Operating Only Alarm System, 1^{st}-principles

There are now three distinct columns, each representing a unique set of circumstances to compare against each other. The first column represents a theoretically optimal simulation, where we should expect that the simulated probability values will converge to the numerically computed ones. The conditions are theoretically optimal because the simulated model is the same as the one used for design, as well as the implementation type (default sub-interval approximation). Also, the alarm systems are continuously operating on a 24-hour basis. As seen in Tables 7.30-7.33, the simulations often need to be run for a longer time in order for the probabilities to converge on the numerically computed values. With the exception of Table 7.31, all of the probability computations based on simulation are at 100%. In Table 7.31, the false alarm counts are high, therefore, the false alarm (Type I) probability agreement is the best out of all probabilities shown.

When we consider the second column, the *real* first-principles based model is used for simulation, although the implementation type is still the same as was used for design. Also, the alarm systems are still continuously operating on a 24-hour basis. When making the simple change from the Gaussian-stationary approximation of the model to the real first-principles based model itself, significant changes occur. All other parameters are exactly the same, even the run-time length. However, as seen, there are still problems with running the simulation long enough due to the low false alarm and missed detection counts for all cases except for Table 7.31, and the Type II missed

detection count for Table 7.33. In these cases, the simulated probabilities diverge wildly from the numerically computed values, especially for the Type II missed detection probability in Table 7.31. The reason for this is clearly due to the fact that the non-stationary non-Gaussian effects cannot be accurately modelled, as we found before and discussed in detail for Tables 7.26- 7.29. When observing the final column, a similar analysis can be made as for the last case, that when we adhere to the schedule given in Table 7.1, the deviations in simulated probabilities climb where there is a sufficient count. Also, we are changing the implementation type, so that lends to the changes as well.

Although the aggregate input disturbance statistics of the first-principles based-model may match those of the Gaussian-stationary approximation of the model on a macro level, the same may not be true on a micro level. In fact, the diurnal periodicity exhibited in the first-principles model-based building temperature time-series data cannot be accurately reproduced by a Gaussian-stationary approximation of the model. Clearly, the same is true for the 2^{nd}-order model based purely on statistics. They do an excellent job at approximating the fundamental spectral characteristics of the aggregation of the data (on a macro scale, such as annual statistics, etc.). However, this appears not to be a sufficient characterization for the periodic nature of the data, especially during start up and shut down times. Yet the most interesting fact to notice is that for the most part, this sub-optimal phenomenon has an advantageous side effect. The missed detection probabilities *decrease*, which is to our advantage from a practical standpoint.

In the implementation of any of the optimal alarm systems presented in this section, it is important to consider that the results shown are open to interpretation. It has been mentioned several times that using the minimax design criteria may result in an apparently very high false alarm and missed detection rate. But recall the technical definition of a false alarm and a missed detection from Chap. 3. If the alarm sounds when the criteria is *not* met, it may be met in the next few time steps. Hence, the alarm just sounded "too early." Similarly, if the alarm does not sound when the criteria *is* met, it may also sound in the next few time steps. Hence the alarm was "too

late." However, the alarms *still* indicate that an upcrossing or an event of some type *will* occur, just not at a given time. Even though our criteria was meant to catch *several* events, not just a single one, as was originally proposed by Svensson [70], timing is still an issue when considering real implementation of the optimal alarm system.

Figures 7.3 and 7.4 illustrate this phenomenon for both the "Svensson model" and the "Federspiel model," respectively. For the "Federspiel model," the building temperature shown in Figure 7.4 is simulated using the basic 2^{nd} order statistical model, not the one introduced in Chap. 6 that is based upon 1^{st} principles. It is obvious that the upcrossings shown *are* predicted by the alarms, but they are not caught exactly at the time dictated by the number of time steps in the prediction window. For example, there are two level crossing events in Fig. 7.4. The first is completely missed throughout the entire prediction window, hence the asterisks indicate missed detections during this period. The second one is not caught at the very beginning of the prediction window, but as it slides forward, notice that there are some correct alarms, followed by a few false alarms. All of the missed detections and false alarms count towards the computation of the probabilities shown in Fig. 7.2. However, on average the heuristic interpretation would yield a much better detection rate. The same logic can be applied to Fig. 7.3, for Svensson's model. In Svensson's original study, when he was considering optimal alarms for a single upcrossing event, the simulated Type I/II error probabilities accurately characterized false alarms and missed detections. However, with our "multiple" complaint prediction window, the Type I/II error probabilities have some flexibility in their interpretation regarding the timing of false alarms and missed detections. Their definitions and the resulting probabilities should therefore not be taken so literally.

7.3 Optimal Control Results

Now that we've discussed all that pertains to optimal alarm systems, it is important to discuss the practical implementation of optimal control, and how it ties to the optimal alarm results.

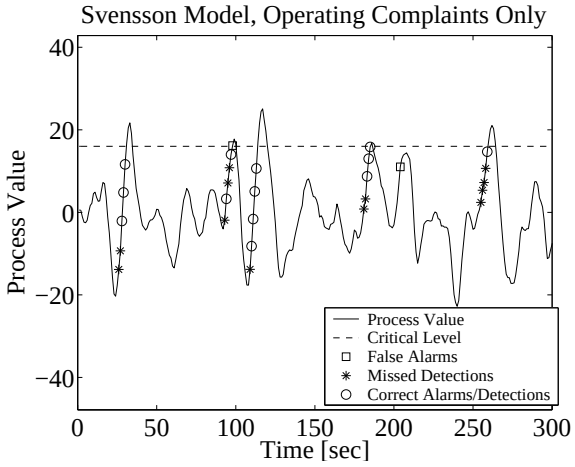

Figure 7.3: Realization of Optimal Alarm System for Svensson's Model

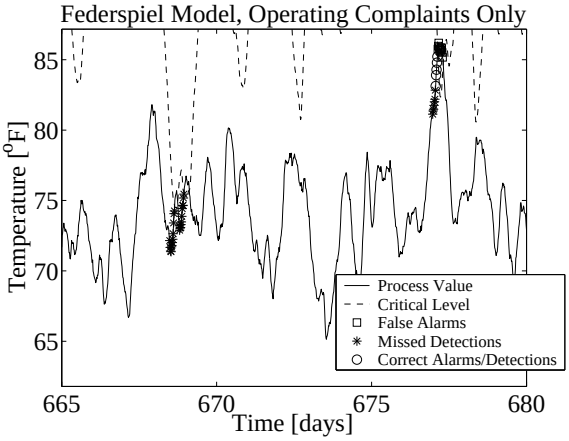

Figure 7.4: Realization of Optimal Alarm System for Federspiel's Model

In the case of response to complaints alone, the optimal control results can clearly stand alone without being tied directly to the optimal alarm results. First we will take a brief look at some basic PI controller issues as far as the "inner loop" is concerned (recall Fig. 6.1).

7.3.1 Inner Loop Controller Details and Event Asymmetry

The PI controller shown in Figs. 1.1 and 6.1 implement simple antiwindup-saturation logic, based upon an exceedance of 50 kW of heating power. For this particular model [46, 47], it was found

that with the incipient disturbances and balance of building thermal inertia, no cooling power is needed in general. Rather, heating power is always needed to maintain the setpoint throughout the year. As such, we need to develop a reasonable actuator saturation limit on this power. The power required is assumed to be practically realized with equipment that will have natural thermodynamic limits on its output capacity. The only bearing that this will have on the results provided in this section is that the setpoint will not be maintained during actuator saturation. This means that the temperature will be lower than the required setpoint, and hence we should expect more cold events in general than hot events. This is clearly the case for the results presented in the previous section for optimal alarm as well. The same PI controller was part of the first-principles based model used to obtain the optimal alarm results in Tables 7.30-7.33 of Sec. 7.2. By observing the statistics in these tables it is evident that cold events greatly outnumber hot events.

However, this is also the case when using real data, or the simple 2^{nd}-order statistics-based model for simulation. The reason for the asymmetry in cold events outnumbering hot events in the case of the 2^{nd}-order statistics-based model is due to the inherent spectral characteristics of the cold complaint process, which has a much higher rate of change ($1.14\frac{^oF}{hr}$ for hot vs. $4.08\frac{^oF}{hr}$ for cold), and a slightly higher variance (5.06^oF for hot vs. 6.14^oF for cold) . The relative means are also different (22.74^oF for hot vs. 17.84^oF for cold), playing a minor role in the asymmetry of the number of events. However, when using the real or modelled data and operating the alarm systems continuously, as discussed in Sec. 7.2, the non-stationarity of the processes on the cold side contribute significantly to this asymmetry as well.

7.3.2 Baseline Setpoint Control Policies

To begin the discussion of optimal control behavior, we first need to review all of the possible control and alarm events that may occur during testing and simulation. The different types of hot and cold events are summarized in Table 7.34. In this table, the optimal operating complaint response shown is triggered by a single upcrossing event, during the “operating complaint period”

defined in Table 7.1. Recall that an upcrossing in simulation is prompted by the following condition: $\{z_{k-1} < 0, z_k > 0\}$. Similarly, when considering the optimal arrival complaint response, we describe the corresponding event as two consecutive exceedances at any point during the arrival period, or as $\{z_{k-1} > 0, z_k > 0\}$. The reason for the definition of the arrival event as two consecutive exceedances is to count only arrival complaints, and not operating complaints. We want building occupants to arrive in a pre-existing complaint condition. It is also possible to define the arrival complaint region by using a single exceedance in lieu of two consecutive exceedances, however, doing so will limit the range of the time in a complaint status upon the arrival of building occupants.

All optimal *alarm* responses shown in Table 7.34 are triggered by a single binary decision that needs to be performed. Recall from Sec. 7.2 that we can use the exact alarm condition to trigger the binary decision. An example is the condition shown as Eqn. 3.2 from Chap. 3, for operating complaints only. This would require the numerical computation of the corresponding integral only. The computational burden would not at all hinder the real-time control performance of the system, because it takes on the order of hundredths of a second to compute.

Event Type	**Possible Event Responses**	**Formulae Required**
Operating Complaint Event	Baseline Operating Complaint Response	Eqns. 7.1-7.2
	Optimal Operating Complaint Response	Eqns. 6.26-6.29, 6.30-6.35
Arrival Complaint Event	Baseline Arrival Complaint Response	Eqns. 7.1-7.2
	Optimal Arrival Complaint Response	Eqns. 6.26-6.29, 6.36-6.41
Alarm Event for Arrival/Operating Complaint Period	Baseline Arrival/Operating Alarm Response	Eqns. 7.1-7.2
	Optimal Arrival/Operating Alarm Response	Eqns. 6.26-6.29, 6.50-6.53
Alarm Event for "Operating Only" Complaint Period	Baseline "Operating Only" Alarm Response	Eqns. 7.1-7.2
	Optimal "Operating Only" Alarm Response	Eqns. 6.26-6.29, 6.44-6.49

Table 7.34: List of Complaint and Alarm Events

All *baseline* complaint responses shown in Table 7.34 are simulated by assuming an adhoc strategy of facility operators responding to events. The response is simulated by changing the magnitude of the thermostat setting by a random amount in the appropriate direction. The random magnitude of the change is governed by a robust regression analysis performed on actual thermostat setting changes logged in a commercial building similar to the candidate buildings identified in Sec. 7.1. The method used was based on a reweighted least-squares regression as in Stone [68], and devised by Federspiel. The adhoc formula governing setpoint changes is shown in Eqns. 7.1-7.2. The parameters 71.0954 and 0.0148 were obtained via the stated robust regression analysis. The additional $1.5\mathcal{Z}$ term is meant to characterize the realistic random nature of typical thermostat setting changes. It is obtained from the standard deviation of the residuals from the reweighted least squares model with an outlier removed.

$$g_k = 71.0954 + 0.0148 r_{k-1} - 1.5\mathcal{Z} \tag{7.1}$$

$$r_k = \begin{cases} s(-g_k) & \text{hot event} \\ s(g_k) & \text{cold event} \end{cases} \tag{7.2}$$

$$\text{where } \mathcal{Z} \sim \mathcal{N}(0,1)$$

$$\text{Recall } s(\delta) = \begin{cases} 1 & \delta > 0 \\ 0 & \delta \leq 0 \end{cases}$$

This baseline setpoint control policy will also be used for comparison to the results of optimal control, later in the section. The common metric for comparison is one that was described previously, where both complaint and energy costs are considered. However, if we consider the case in which the response to complaint or alarm events is automated, or "semi-automated," there is no need to dispatch a facility technician, and therefore no labor charges are incurred. Therefore the complaint cost as defined is eliminated via automation, but it may still be of use to quantify this same metric in a different way to quantify thermal discomfort. It may be assumed that shifting this cost to quantify discomfort rather than complaint costs is akin to attaching a dollar value to

discomfort as investigated by Mozer et.al. [52, 53].

Therefore the actual cost metric is as in Eqn. 2.5 where $J = \sum_{k=0}^{L} R_k P_k T_s + C_k$, even though the cost function penalty is actually the square cost of energy *conditioned* on the type of event being responded to, as in Eqn. 6.23, $J = T_s^2 E[R_N^2 h_N^2 + \sum_{k=0}^{N-1} R_k^2 h_k^2 | \mathcal{E}]$. Pragmatically, our interest is in the former cost function, even though it is the latter that is used to derive the optimal control law. Therefore, the results presented are not expected to be theoretically optimal, since we are using a metric for comparison that has not actually been used to derive the optimal control law. In fact, they will be sub-optimal, and perhaps they will exceed the costs of the baseline strategy over a portion of the range of a given modelling parameter.

There are actually two different baseline setpoint policies that we'd like to consider. One represents *no* response to *any* complaint events, where the setpoints remain at nominal statically optimized thermostat settings. This optimization was described in Sec. 6.1, but it does not consider energy costs as demonstrated by Federspiel [19, 21, 22]. The second baseline setpoint policy is the one we've described earlier with Eqns. 7.1-7.2 and presented in Table 7.34. A realization of the static optimization control policy and its counterpart control actuation and price signal are provided in Fig. 7.5.

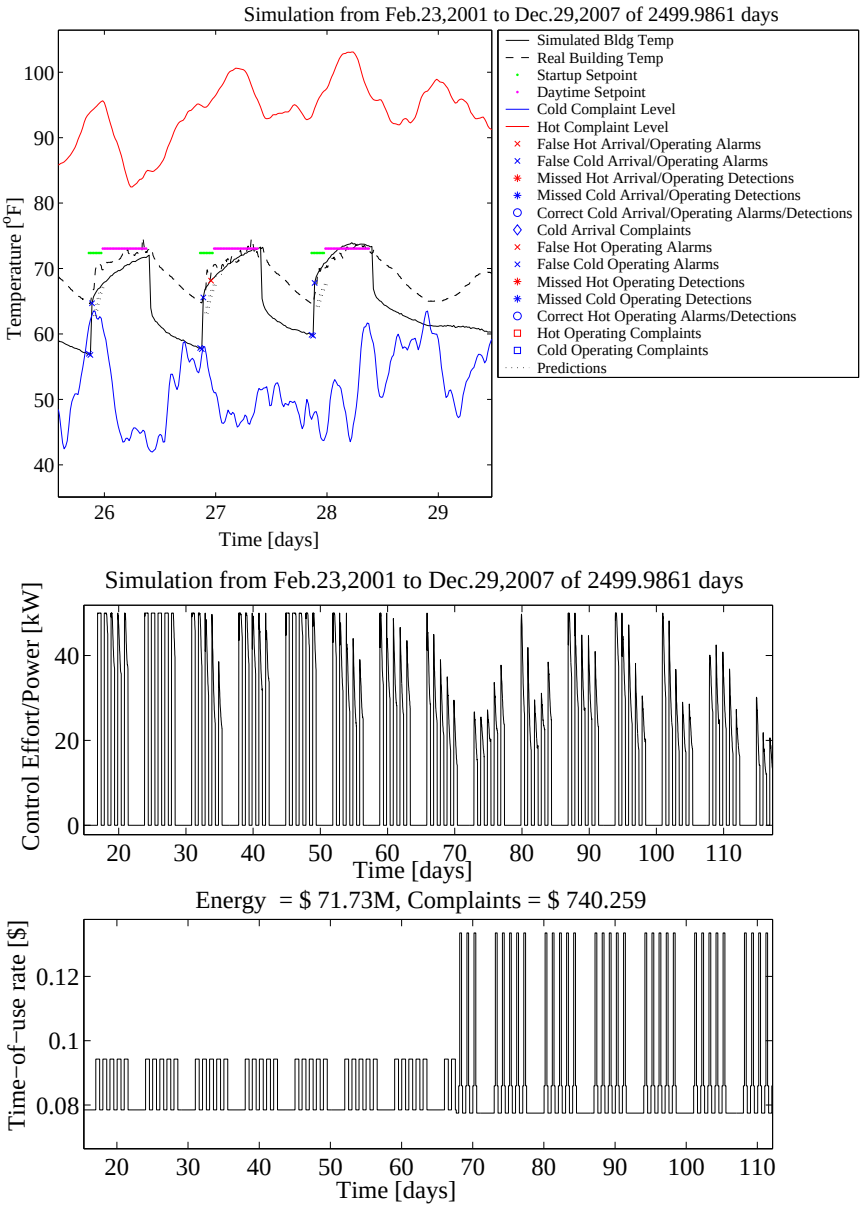

Figure 7.5: Nominal Statically Optimized Setpoint Control Policy

The top graph in Fig. 7.5 shows an actual realization with all processes and events that occur during the simulation. Although the simulation run time is shown to be 2500 days, only a very short slice of simulation time is actually displayed, from 26 to 29 days. Therefore, not all of the events shown on the legend appear within this timeframe. However, some fundamental

characteristics of the plot can be pointed out. The red and blue lines represent the hot and cold complaint processes, respectively. The solid black line represents the building temperature process according to the first-principles based model and corresponding schedule, and the dashed black line represents real building temperature data from the prototype zone. For the purposes of the study in this subsection, all complaint events are taken as interactions of the model-based solid black line building temperature with the complaint level processes.

Although none appear in the graph for the timeframe shown, any red or blue squares ($\square$) represent hot or cold or operating complaints, respectively. Also, any red or blue diamonds ($\diamondsuit$) represent hot or cold or arrival complaints, respectively. The events that do appear in this short time frame are hot and cold false arrival/operating alarms, represented by the red and blue cross ($\times$) indicators. No other types of alarm indicators (for missed detections, or hot/cold operating alarms) appear in the graph. Another important association of the false alarms shown in the graph are the predictions appearing as dotted black lines that correspond to the alarms shown. Finally, the green flat lines represent the statically optimized arrival complaint reference thermostat setting for the arrival/operating alarm period (4:30 am to 7:30 am). The magenta flat lines represent the nominal thermostat setting for the operating alarm period (7:30 am - 5:00 pm, also statically optimized).

On the lower portion of Fig. 7.5, we can see the graphs for the control power as a function of time, as well as the price of electricity as it varies with the time of day and season. Notice that the actuator saturates at 50 kW several times in the simulation period shown, where the antiwindup logic kicks in. The seasonal variation of the electricity price can also be observed in the plot, and occurs at around day 68. The plot also shows the amount of money spent on energy and complaints for the entire duration of the 2500 day simulation period, even though just under 100 days are shown. A similar realization of the *baseline* setpoint control policy is provided in Fig. 7.6.

The main feature of the graph shown in Fig. 7.6 illustrates a hot operating complaint, at day 109, which is responded to by using the baseline setpoint control policy. Notice the dip in the building temperature around the middle of the operating period during the day, as the thermostat

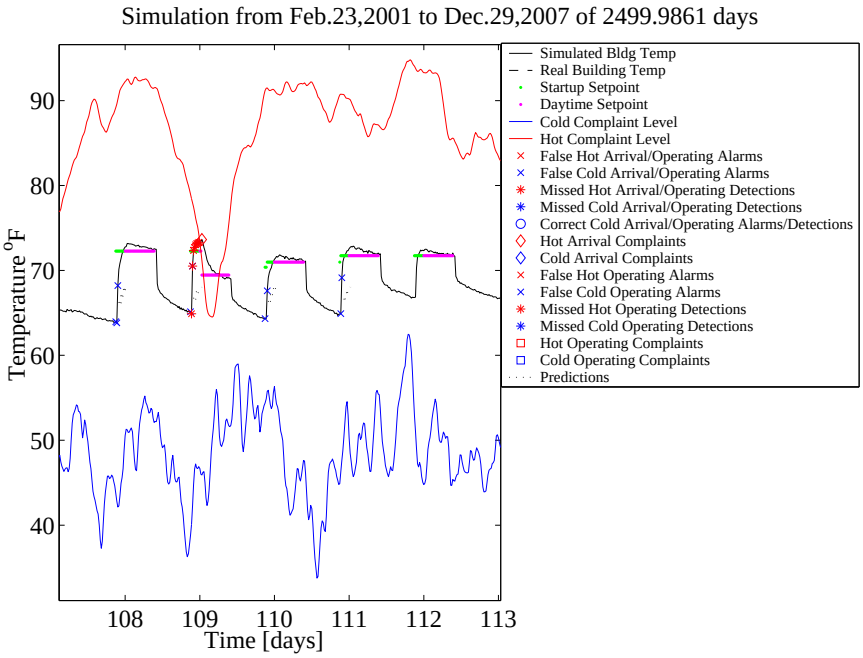

Figure 7.6: Baseline Setpoint Control Policy

setting is changed from 72.27 ^oF to 69.45 ^oF. There are also a few subsequent minor adjustments made due to a few false cold arrival alarms in the next 2 days. We forgo the repetitive displays of the counterpart control actuation and price signal graphs, because qualitatively they are very similar to those shown in Fig. 7.5. The most important information gleaned from these two omitted graphs are the metrics quantifying the energy and complaint (discomfort) costs. They will be summarized along with the static results in Table 7.35 shortly.

Fig. 7.5 illustrates the costs for energy, complaints (discomfort) and the sum of both, which were recorded during a 2500 day simulation (6.85 years). The energy computation was performed according the rate scenario used by Federspiel in his study [21, 22], the E-19 electrical energy tariff offered by PG&E. Also, the complaint cost, C_k, is proportional to the number of hours charged. Let $C_k = HLN_k$, where H is the average number of hours charged in response to a thermal sensation complaint, L is the total hourly labor rate for technicians who handle complaints, and N_k is a

discrete binary variable characterizing the "complaint status" at time k, with a value of 1 for a complaint and 0 for no complaint. According to studies by Federspiel [20, 21], $L = 1.85$ hours, and the total hourly wage for an HVAC mechanic in Sacramento, California, is $L = \$30.78$/hour.

The results of the simulations are summarized in Table 7.35. Notice that the statically optimized policy (not responding to any complaints) has a higher cost for both energy and complaints (discomfort) than when using the baseline optimal control policy, as expected. These results will also be compared to the results for automated optimal control and alarm implementation using the same simulation run-time in the next subsection.

Setpoint Policy	**Energy Cost**	**Complaint (Discomfort) Cost**	**Total Cost**
Nominal Statically Optimized Setpoint Control Policy	\$71,727,319.85	\$740.259	\$71,728,060.11
Baseline Setpoint Control Policy	\$69,970,848.12	\$455.54	\$69,971,303.66

Table 7.35: Table of Baseline and Static Setpoint Policy Metrics

We are interested in an assessment of the sub-optimality of using Eqn. 2.5 where $J = \sum_{k=0}^{L} R_k P_k T_s + C_k$ as a metric, even though the cost function penalty is actually the square cost of energy *conditioned* on the type of event being responded to, as in Eqn. 6.23, $J = T_s^2 E[R_N^2 h_N^2 + \sum_{k=0}^{N-1} R_k^2 h_k^2 | \mathcal{E}]$. It will be determined if the costs associated with the baseline and static setpoint control strategies exceed that of the metric being measured in simulation, over a portion of the

range, or the *entire* range of a given modelling parameter.

7.3.3 Optimal Control Policy

The modelling parameter that has been alluded to earlier as an instrument to examine the optimal control policy and compare to the other baseline control strategies is naturally the border probability, P_b. Changing the border probability will result in greatly different responses to alarms. The selection of an optimal P_b value is a design criteria that affects the sensitivity of building operation with respect to false alarms and missed detections. However, it can also be looked at parametrically in terms of the subsequent automated thermostat setting control response. In fact, it is instructive to consider the "average responses," as shown in Figs. 6.3 and 6.4, where the optimal control, u_k, is plotted over the desired finite horizon, representing the deviation from the nominal thermostat setting, $\bar{r}$. Therefore, in Figs. 7.7 and 7.8, the average responses to alarms for several different values of P_b are shown. The conditions for which the control solutions were computed are exactly that same as those used for Figs. 6.3 and 6.4. Responses to complaints cannot be parameterized by P_b, because doing so is unique to alarms, and we are trying to determine the utility of the alarm-based control, as opposed to simple complaint response.

In Figs. 7.7 and 7.8, there are 6 different values for P_b shown in each graph. In Fig. 7.7, for hot alarms, all 6 different values of P_b can clearly be distinguished. The corresponding P_b values come from the following set: $\{0.0001, 0.0002, 0.001, 0.002, 0.00536, 0.01\}$. For the alarm-based response to operating complaints on the top of Fig. 7.7, there are some extreme responses. This was also evident in Chap. 6. Here we see that as P_b increases, so does the maximum magnitude of the response. This illustrates the sensitivity of the control algorithm to the values of P_b, because of the small range of values that P_b is plotted for.

If we select a very small value for $P_b = 0.0001$, i.e. the smallest one from the list shown, then the maximum magnitude of the control response over the duration of the finite horizon is $\approx 27^o F$. This value is reasonable, because it will decrease the building temperature according to

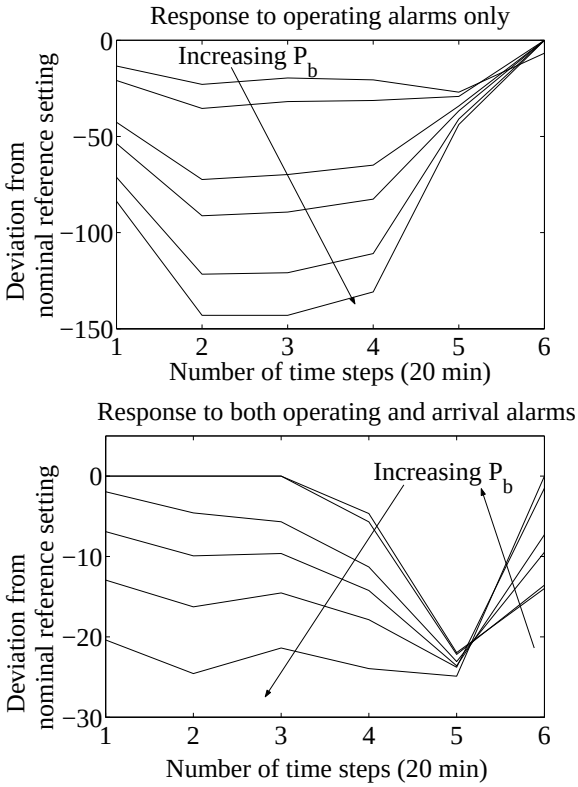

Figure 7.7: Optimal Alarm-Based Control Policy for Hot Alarms (Different P_b values)

the magnitude of the response required by alarm system's "tuning." However, as the alarm system's tuning changes, and P_b increases a few orders of magnitude, the alarm system response is to decrease the thermostat setting more and more. Consequently, the maximum magnitude of the thermostat setting response peaks at 143^oF. This, of course, seems to be an unreasonable change from a thermal comfort standpoint. However, consider the fact that the optimal control cost function is truly based on penalizing energy, as well as conditioned on the alarm statistics. As such, decreasing the thermostat setting significantly will satisfy the dual objective of reducing the temperature in response to a potential hot complaint, and reducing the energy consumed in doing so as well.

Furthermore, because of the natural slow dynamics of the closed-loop system, the building will take some time to respond to the change in a thermostat setting. These dynamics are taken

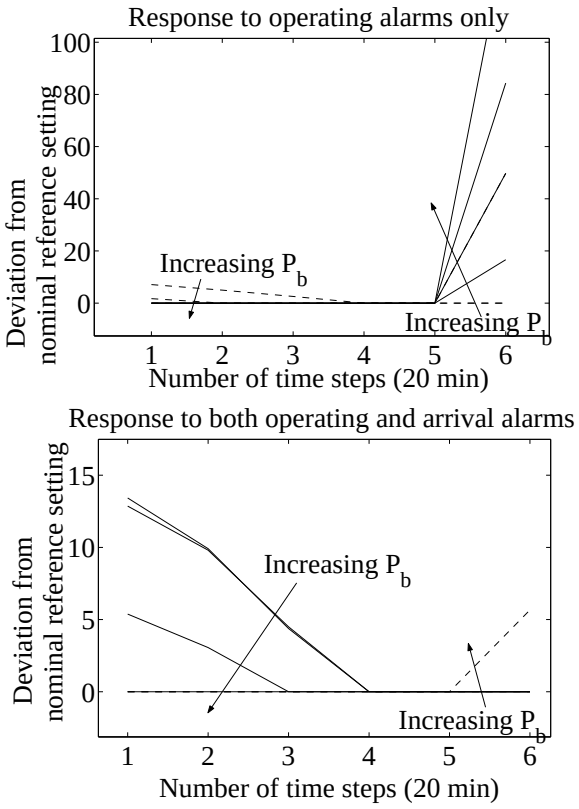

Figure 7.8: Optimal Alarm-Based Control Policy for Cold Alarms (Different P_b values)

into account in the derivation and execution of the control law. As such, the maximum magnitude of the thermostat setting does not provide a true quantitative measure of the change in building temperature that will take place. In fact, the change in the building temperature over the plotted horizon will be mitigated by the fact that the window is finite, and the magnitude of the response decreases considerably over this horizon. However, we must still be very careful about which method we use to design the optimal alarm system. Obviously, the control response is very sensitive to the P_b value selected according to a design criterion that may not accurately represent the true costs that need to be penalized. Recall that a thorough discussion of assigning weights in formation of an optimal alarm system design cost function was discussed in Chap. 5. This discussion may provide a viable alternative to the minimax design criterion used for the examples in this chapter.

Another factor influencing the alarm-based control system is the fact that the multi-dimensional method was used to approximate the exact alarm region, as discussed in Chap. 6. It may also be the case that the method of implementation of the multi-dimensional approximation with the control algorithm is sub-optimal. Recall that as currently implemented, the alarm region is split up into pieces, instead of treated as a whole. Although the multi-dimensional approximation of the exact alarm region spans all of the time slices in the horizon, it may still be feasible to attempt to implement the control law presented in Sec. 6.2.4, using the entire approximation on a step-by-step basis.

On the bottom of Fig. 7.7, the alarm-based response that considers both arrival and operating complaints is illustrated. Evidently, much of the same behavior exist as for the case above it. One of the main differences is in the maximum magnitudes of the control responses to these types of alarms. This difference is clearly influenced by the implementation of the approximated alarm region, which is distinct from the previous one. Also, the qualitative nature of the responses are different in that if P_b is small enough, there is no immediate control response, and larger magnitudes occur later in the control horizon. This may also be influenced by the nature of the different region which accounts for arrivals, as well as the fact that there is a $d = 1$ hour look-ahead prediction window, in which the rate of electricity changes according to Fig. 6.5.

In Fig. 7.8, for cold alarms, all 6 different values of P_b *cannot* clearly be distinguished. The corresponding P_b values come from the following set: $\{0.001, 0.002, 0.01, 0.02, 0.027867, 0.1\}$. On the top graph, when responding to cold operating alarms only, as P_b increases, the qualitative nature of the response changes significantly. For the smallest two values of P_b, shown as dotted lines, the response gradually decreases until it dies off to 0 at the end of the control horizon. For the 4 greatest values of P_b, shown as solid lines, the opposite happens, where the response is 0 until the very end of the control horizon, where it increases proportionally with the value of P_b. A similar qualitative description characterizes the behavior of the alarm response for both arrival and operating complaints on the bottom graph. However, in this case, the smallest *three* values of P_b

are shown as *solid* lines, and the response gradually decreases until it dies off to 0 at the end of the control horizon. Interestingly, there is no response at all the for alarms corresponding to the next larger 2 values of P_b=0.02,0.027867. Finally, for the largest value of P_b, shown as dotted line, the response is 0 until the very end of the control horizon, where it increases. The magnitudes of all of the responses to the cold alarms shown in Fig. 7.8 are reasonable, with the possible exception of the higher P_b values for the "operating only" alarm system. Here, the same explanation as was used to justify the extreme results for Fig. 7.7 applies.

For the results in Figs. 7.7 and 7.8, there are cases in which there is no immediate response to the alarm or in some cases no response at all. This can be attributed to the fact that the control was posed as a constrained input problem, invoking the application of Pontryagin's Minimum Principle to arrive at a solution. The hard constraints on the control problem disallow any part of the control trajectory into the inadmissible region. As such, the "bulk" of the response is often "saved for later," at times when, for example, the price of electricity is more conducive to an admissible control action that favors reduced energy usage. The same is true of the response to complaints as opposed to alarm, as shown in the top 2 graphs of Fig. 6.3 in Chap. 6.

Now that the importance of the sensitivity of the control response to alarm design has been established, it is of interest to compare the optimal control policy to the other baseline control strategies using the border probability, P_b. Alarm system response was executed for the baseline strategies as well. We can compare the relative magnitude of the results for the baseline control strategies shown in Table 7.35 by superimposing them on the same plot. Recall that we are interested in an assessment of the sub-optimality of Eqn. 2.5 where $J = \sum_{k=0}^{L} R_k P_k T_s + C_k$. We will now investigate the costs of the baseline and static setpoint control strategies. It will be determined if they exceed the same metric of the optimal control strategy being measured in simulation over a portion of the range, or the *entire* range of P_b values under consideration. This is an important metric even though we are automating the strategy (i.e. C_k is typically used for unautomated operation), and it is not the actual cost function used to derived the control law. The complaint cost portion of it

is important from an economic standpoint, and good a indicator of thermal discomfort nonetheless. An unintended additional feature of the comparison is the quantification of the improved optimal response to *complaints* over the baseline response, in addition to alarms.

Figs. 7.9-7.11 illustrate a decomposition of the metric described above, with the baseline and static data from Table 7.35 superimposed for comparison. The P_b values shown represents the *average* values for all alarm systems in operation for that simulation. This includes all 4 alarm systems: hot and cold alarm systems and both operating only and arrival/operating alarm systems. The simulation run time of 2500 days (6.85 years) was used for all data points shown in Figs. 7.9-7.11.

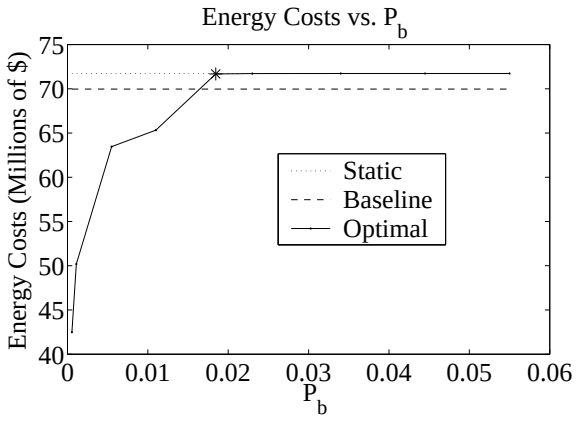

Figure 7.9: Energy Costs for Different P_b values

Notice that the optimal average value of P_b corresponding to the minimax design criteria used for optimal alarm design (0.01845) yields an optimal energy cost of \$71.67M and an optimal complaint (discomfort) cost of \$170.829. Both of these points are shown with asterisks in Figs. 7.9 and 7.10, respectively. This appears to be *close* to the best tradeoff between both energy and complaint (discomfort) costs, as well as the *best* tradeoff between false alarms and missed detections. Comparing sub-optimal energy costs to the baseline and static energy costs shown in Fig. 7.9, we see that the general trend is for the static cost to act as an upper bound on the optimal energy costs.

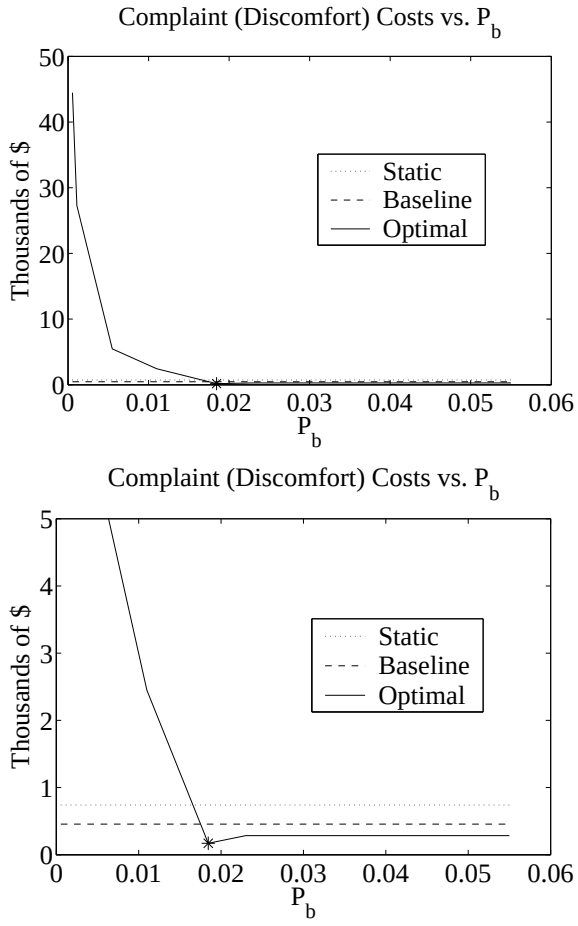

Figure 7.10: Complaint (Discomfort) Costs for Different P_b values

However, the baseline cost is exceeded by the sub-optimal cost just to the left of the minimax-based optimal P_b value. Therefore, we now have evidence of the sub-optimality of using a different cost function for comparison than was actually used to derive the control law. The same phenomenon holds true in general for the total cost, as shown in Fig. 7.11. Therefore, we can make a reasonable conclusion pertaining to the sub-optimality we intended to investigate. The resulting suboptimal energy costs will always be lower than the *static* energy costs for all values of P_b. If we choose $P_b < 0.0165$, where the suboptimal curve intersects the baseline value, the resulting sub-optimal energy costs will also be lower than the *baseline* energy costs.

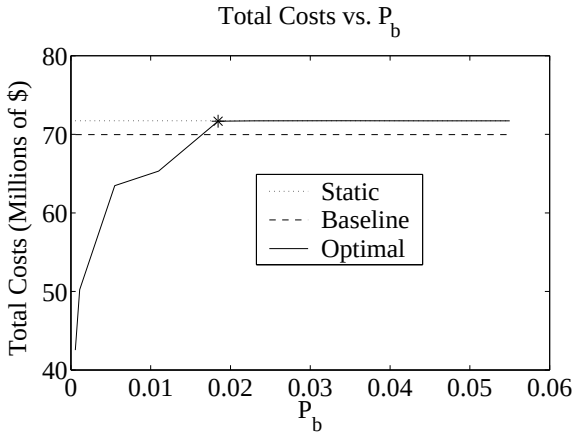

Figure 7.11: Total Costs for Different P_b values

Now if we compare sub-optimal complaint or discomfort costs to the corresponding baseline and static costs shown in Fig. 7.10, we see the opposite trend. Here, the sub-optimal complaint/discomfort cost starts off very high for low values of P_b, but as it is increased, the sub-optimal cost decreases rapidly below the respective baseline and static costs. Two graphs are shown to indicate the qualitative nature of this more clearly. The top graph shows all plotted data points, and the bottom graph zooms in on the area of interest. Evidence of sub-optimality is still present due to the exceedance of the static and baseline costs for small values of P_b. If we choose $P_b > 0.0166$, where the suboptimal curve intersects the static value, the resulting sub-optimal complaint/discomfort costs will be lower than the static complaint/discomfort costs. Similarly, if we choose $P_b > 0.0175$, where the suboptimal curve intersects the baseline value, the resulting sub-optimal complaint/discomfort costs will be lower than the baseline complaint/discomfort costs.

In lieu of using the total cost metric $J = \sum_{k=0}^{L} R_k P_k T_s + C_k$, we may try to apply a multi-objective criteria for optimality, such that both sub-optimal energy costs ($J = \sum_{k=0}^{L} R_k P_k T_s$) and complaint costs ($J = \sum_{k=0}^{L} C_k$) fall below the corresponding baseline costs. However, the intersection of the two resulting solution regions, $P_b < 0.0165$ and $P_b > 0.0175$, is empty. This may not always necessarily be the case due to inherent statistical error involved in the results, since they

are all based on simulation. Certainly if we choose $P_b > 0.0175$, the sub-optimal costs are always lower than both of the corresponding *static* costs. But we are more interested in obtaining a sub-optimal control that performs better than the *baseline* strategy, because that is our benchmark for how facility operators currently respond to complaints in buildings. When considering the original metric quantifying total cost from Fig. 7.11, recall that the same phenomenon for the energy costs alone hold true. That is, if we choose $P_b < 0.0165$, the resulting total cost will always be lower than the total baseline or static cost. However, there is no unique solution in this case, which is less valuable to designers of the alarm-based control system. It is possible to consider alternative methods for finding a solution to the sub-optimal control, such that an optimal P_b can be found to satisfy a reasonable tradeoff between energy and comfort considerations that beats the baseline costs.

The best tradeoff between both energy and complaint (discomfort) costs may be a variant of the minimax design criterion. This is true only if we consider the costs to be weighted, as in Martin et. al. [48], rather than on an equal footing with each other. Doing the latter will default to the non-unique complaint/discomfort cost solution of $P_b > 0.0175$. The choice of these weights can be left purely to the discretion of the designer, but in order to arrive at a unique solution that provides a total weighted cost that is less than the baseline costs, the weights must be chosen carefully. For example, if we weight the complaint cost at least 3 orders of magnitude greater (i.e 1000 times greater), and keep the weight on the energy cost at unity, we can arrive at a unique solution of a very small value for P_b. The energy cost should never be weighted because we know these costs are inflexible, and we always want to compare it to the benchmark baseline cost having the same order of magnitude. However, weighting the complaint costs is perfectly acceptable considering the fact that the complaint cost has an alternative interpretation as a discomfort cost, which is more loosely defined.

As such, if a multi-objective criteria for optimality will not yield a non-empty solution, or using the non-unique total cost metric solution is undesirable to the designer, it is practical to use

this weighted variant of the minimax formulation. Using any of these alternatives will result in an economical, energy efficient optimal alarm-based control system whose implementation falls below baseline costs. It will also provide the most correct, robust and adequate preventative and corrective response to thermal discomfort events. In fact, it is possible to choose a weight that will result in the same solution corresponding to the minimax design criteria used for optimal alarm design, although it will not necessarily result in any economical advantage over the baseline costs. However, it has been the method for presenting all of the results in this chapter thus far. Therefore, a realization of the corresponding optimal control thermostat setting policy and its counterpart control actuation and price signals are provided in Fig. 7.12.

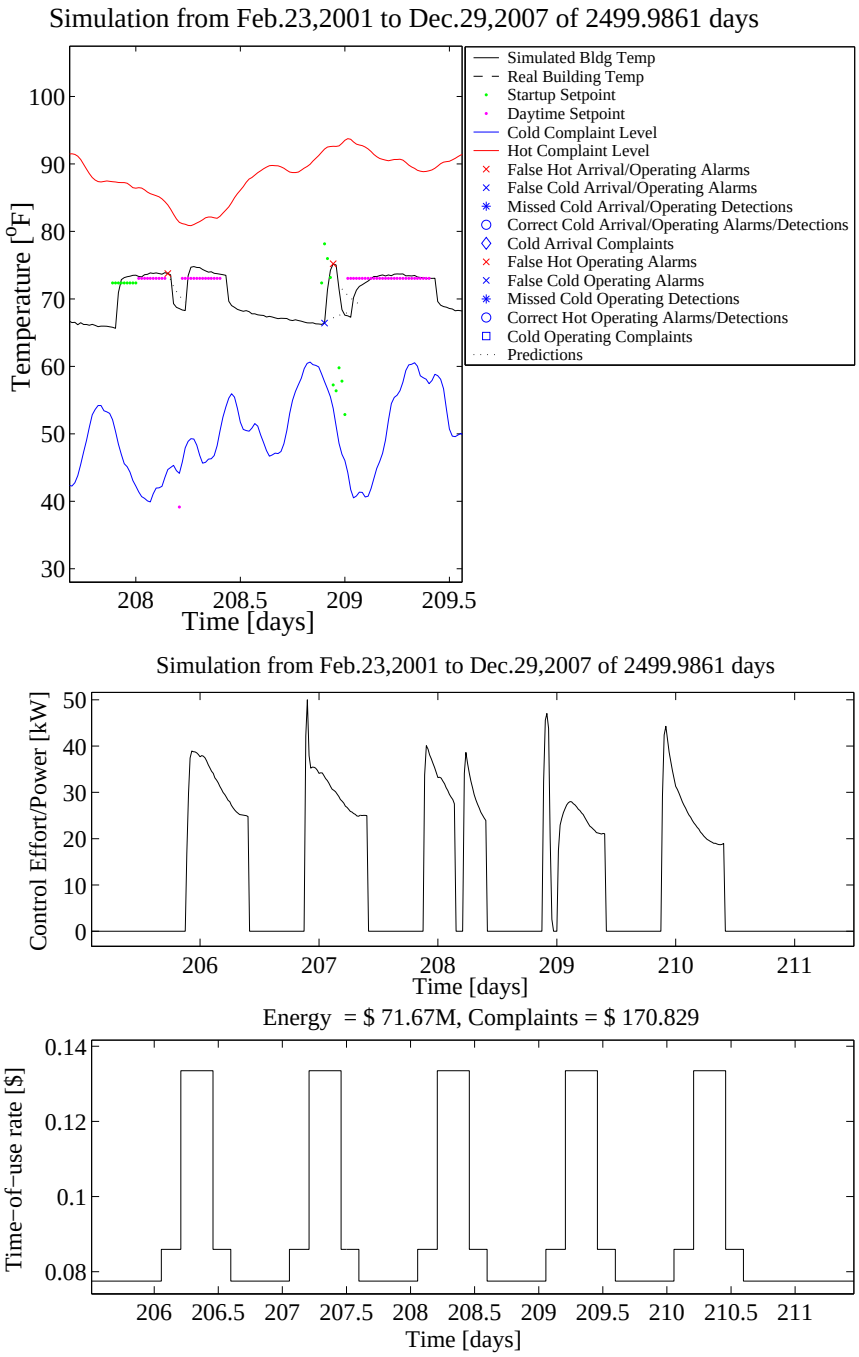

Figure 7.12: Optimal Thermostat Setting Control Policy for Minimax-Designed Alarm System

Notice that in Fig. 7.12, although 2500 days in total were simulated, only two full days are shown in the top graph. The legend and major features are very similar to the explanations provided for Figs. 7.5 and 7.6. There are three basic events that take place during the time period shown. For all other periods, the thermostat settings are scheduled according to the statically optimized setpoints. All of the events shown are false alarms. The first alarm event occurs during operating hours, and the response is evident in the subsequent thermostat setting change. Only the last setpoint close to 40^oF in the control horizon appears on the graph, because the remaining optimal setpoints fall far below the lower limit of the graph. The lowest optimal point is close to -50^oF. However, notice the response to the thermostat setting causes the building temperature excursion to dip down only to 68^oF. This is an example of the phenomenon mentioned earlier, where the building will take some time to respond to the change in a thermostat setting because of the natural slow dynamics of the closed-loop system. Also, the change in the building temperature over the plotted horizon is mitigated by the fact that the window is finite.

For the remaining two false alarm events that occur the next day, a similar excursion in building temperature results, during the arrival period. Here, a cold false alarm is immediately followed by a hot false alarm, due to the low temperature at the beginning of the arrival period, followed by the start-up of the HVAC system that cause the temperature to increase very rapidly. The subsequent control response is for the thermostat setting to change according to the perceived thermal comfort demands, and the building temperature to follow. Notice the control power used during these two days when making these responses, shown in the second graph below. Energy is saved by responding to the events in the manner they are, even though they are false alarms and no actual complaint events occur. The perception of thermal discomfort will not always be perfectly predicted due to the nature of the alarm systems' inherently imperfect false alarm/missed detection tradeoffs. Yet at times when the price of electricity is at a premium, according to the last graph in Fig. 7.12, the energy savings during this time may be worth the temporary discomfort of building occupants. In the long run, the energy savings reaped by these false alarms will balance against

times when the alarms will actually be correct, and both thermal comfort and energy objectives are simultaneously satisfied.

7.3.4 "Alarm-Based Control" System Utility

Recall the discussion from Sec. 7.2 on the utility of the alarm system when designed for an unautomated response. If we consider the case in which the response is automated, the utility of the optimal alarm system changes, and energy costs have implications on optimal control. The reason that the utility of the alarm system must be tied to energy considerations is because complaint costs, C_k, as defined with Eqn. 2.5 where $J = \sum_{k=0}^{L} R_k P_k T_s + C_k$, have been eliminated via automation. That is why it was necessary for us to shift this cost to quantify discomfort rather than complaints. As such, we now investigate the energy cost implications of using an optimal alarm system.

In general, for response to any type of complaints, whether the result of missed detections or not, and response to any type of alarms, *comfort costs energy.* Any response to an event, whether anticipated or caught after the fact without prior warning will be to change the setpoint. This clearly incurs some sort of energy cost, whether positive or negative. However, if we consider the case of correct alarms/detections, we may incur additional energy costs that we wouldn't otherwise if we hadn't pre-empted the complaint. In other words, if an alarm correctly predicts a complaint, there are two responses, one in response to the alarm, and a subsequent one in response to the actual predicted complaint. Therefore, energy costs will be incurred for both responses.

If we consider the details for the case of correct alarms/detections, the magnitude of energy used in response to the alarm and complaint events depend on several factors. The first factor is whether the event is a hot or a cold event, the second factor is whether the zone requires heating or cooling power (i.e. a seasonal basis), and the final factor has to do with the overall statistics of hot events vs. cold events. If the response is to a hot event during "heating season," then negative setpoint changes will reduce the heating power used, and it's favorable in terms of both thermal comfort and energy objectives. The same holds true (reduction in cooling power) if the response is

to a cold event during "cooling season." However, if hot events take place during cooling season, or cold events occur during heating season, then *additional* costs are incurred due to energy usage in response to building thermal comfort. Changing the setpoint in response to events in such a manner demands more power usage.

On an aggregate (i.e. annual) level, if on average the number of hot events is the same as the number of cold events, regardless if it is heating or cooling season, the reductions and increases in energy will balance each other out. However, if the statistics of hot and cold events are asymmetric, then on a seasonal basis, the energy ramifications will either be favorable or unfavorable. For example, during heating season, if there are more hot than cold events, then annual energy usage should be favorable. However, if there are more cold than hot events, annual energy usage is unfavorable. Likewise, during cooling season, if there are more hot than cold events, the annual energy usage is unfavorable, otherwise it is favorable. Therefore we see that asymmetry is favorable with respect to energy usage when there are more events of a type that are atypical of the season. Symmetry in general is neither favorable nor unfavorable with respect to aggregate energy usage.

If we consider the simulation results presented in this and the previous section, recall that for this particular model [46, 47], heating power is always needed to maintain the setpoint throughout the year. Furthermore, we've found that in general, there is an asymmetry in cold events outnumbering hot events. As a result, overall there is an unfavorable conclusion with regard to energy. Even though Fig. 7.12 demonstrates more hot than cold events, and energy is saved by responding to the events in the manner they are, it will still be the case that cold events outnumber hot events on average.

However, we have only considered the energy cost part of the discussion thus far. If we don't consider quantifying discomfort, then on average *correct* alarms will *always* incur additional charges. As stated several times previously, because complaint costs are assumed to be eliminated with automation of the response to system events, we *must* consider quantifying the cost of discomfort somehow, and doing so will counteract the resulting unfavorable energy implications for correct

alarms. If we consider false alarms and missed detections, we see that they will also always cost *extra* due to the stated energy considerations. Unfortunately, there is no counteracting thermal comfort advantage in the case of false alarms, because no actual complaints occur. In fact, there is most likely a thermal comfort disadvantage that results from response to false alarms when there is no pending complaint condition. The response is not expected, and may tend to push the thermal discomfort envelope in the opposite direction than was originally desired. Missed detections, however, will certainly cost extra overall with respect to energy, but the counterbalance of directly responding to thermal sensation complaints presents a clear comfort benefit.

The conclusions thus far concerning energy have been based on aggregate statistics, and as evidenced in Fig. 7.12, event responses will not *always* result in unfavorable energy excursions. Remember also that the conclusions stated here are only valid for a model of the prototype building whose accuracy in quantifying and implementing actual energy usage is dubious at best. We would expect that in a realistic scenario, there is actually both a cooling and a heating season, and the hot and cold statistics would be unique to that particular zone. Therefore, the conclusions to be made in a realistic scenario may differ greatly from the results presented here for our prototype.

It is also important to remember that there is another avenue of control for determining the energy outcome, whether it is a simulated or a realistic scenario. The optimal alarm design procedure can be tailored to account for energy considerations. The optimal value of P_b can be determined by creating a energy-based cost function which uses the Type I/II error probabilities. For example, in the results for the model of the prototype, the energy implications were unfavorable overall. As such, we can assign weighting to the Type I/II error and alarm probabilities that are a function of the amount of energy required to heat or cool that zone. Therefore the size and thermal properties of the zone may be important in establishing these weights as well. The resulting solution to the design problem would be to find the optimal P_b that mitigates unfavorable energy conditions. We've seen a very similar presentation of how P_b can be chosen to favor reduced energy usage earlier in the section, from the discussion of Fig. 7.9.

Let us now hypothetically assume that there is no penalty on discomfort, because complaints can all be accurately predicted and prevented by the alarm system. In this case, the alarm system must be designed for a very low missed detection rate, or a very low P_b value. In a sense we are assuming that the alarm system is perfect and all thermal sensation complaints can be perfectly predicted and prevented. Therefore, any complaints that occur *must* be due to some other reason than discomfort (i.e. a fault in the system, problem with an HVAC component, etc.). As a result, any actual missed detections that occur signal that something else is wrong, and on this very extreme side of the spectrum, our *optimal alarm-based control system* has now become a *fault detection system*.

However, there is no perfect optimal alarm system, and all complaints due to discomfort can never be 100% perfectly predicted are prevented by the alarm-based control system. The main reason for this, of course, is that there will always be a trade-off between false alarms and missed detections made during the alarm design stage. This design can take place with numerous considerations in place, as we've seen. Considering the current extreme scenario, where missed detections are highly penalized, many false alarms will be elicited, and result in subsequent automated responses. As such, any actual complaint is most likely due to a fault in the mechanical system, and will require the dispatch of a facility technician. Therefore, no money is saved when choosing a very low value of P_b, because of the dispatch required, even though doing so will result in reduced energy costs overall. The costs and benefits involved with operating the system in a fault detection role for very low P_b values are as follows:

1. High labor dollars due to dispatched facility technicians investigating possible faults.
2. Low energy costs.
3. No missed detections.

On the other end of the spectrum, if false alarms are highly penalized, the alarm system completely loses its utility as a fault detection system, and is more of an optimal alarm-based control

system. The costs involved with switching to an optimal alarm-based control system role for higher values of P_b are as follows:

1. No labor dollars for any reason due to automated responses to all system issues.
2. Higher overall energy costs due to the increased P_b value, and a greater discomfort prevention role.
3. Fewer false alarms.

There is an entire spectrum that spans the different roles that can be played by the system for automated use. Therefore, its operation will save money and benefit building occupants in different ways for different objectives that need to be achieved according to the desires of the user. We see that on one end, the system is a *fault detection system* for low values of P_b, and on the other end, for higher values of P_b, it is an *optimal alarm-based control system.* As such, P_b acts as a tuning parameter that the designer must use to control the objectives that are most important for the operation of the building under their charge.

Chapter 8

Conclusions and Future Research

8.1 Optimal Alarm

To conclude the results on the optimal alarm system for unautomated operation, we revisit the following main points of Chap 7:

- There is no inherent utility in the optimal alarm system, under the following conditions:
  1. The response to complaint or alarm events is unautomated.
  2. Complaint cost is quantified by the labor dollars required to dispatch a facility technician in response to such an event.
  3. A response is mandatory and takes place without regard to energy implications.
- In this case, false alarms must be minimized in order to prevent the exacerbation of the inherent negative utility of the optimal alarm system, by choosing an optimal P_b to maximize $E[C|A]$. However, the alarm will rarely sound by using this criterion, so using the $E[C|A]$ design criterion does not prove to be useful.
- To allow for practical use, but yet for response to complaint or alarm events to be unautomated, the last two conditions stated above must be relaxed.

1. Namely, energy implications *must* be considered, therefore missed detections will be balanced more evenly with false alarms in the selection of P_b by using the minimax design criteria.
2. Furthermore, the utility of the alarm system will have to be quantified differently. It will necessitate a dependence on a dollar value that can be attached to discomfort (as in Mozer et.al. [52, 53]) without requiring a tie to the complaint cost, or response to the actual event by dispatch of a facility technician.

- Monetary quantification of the less tangible thermal discomfort metric and its integration with optimal alarm systems needs further study, even though the complaint cost quantification was more preferable due to its more tangible nature. However, the lack of utility of the standalone optimal alarm system when using complaint costs to measure it necessitates this reformulation.
- Sub-optimality of the results due to assumptions of Gaussianity and stationarity has an advantageous side effect. The missed detection probabilities *decrease*, which is to our advantage from a practical standpoint.
- Due to technicalities, Type I/II error probabilities have some flexibility in their interpretation regarding the timing of false alarms and missed detections. Their definitions and the resulting probabilities should on average be determined heuristically. This interpretation would yield a much more realistic detection rate.

If we consider the relaxed conditions stated in the list above to be implemented, and for the alarm system to have a positive utility for unautomated responses, then the optimal alarm design will be based on the model of empirically obtained 2^{nd} order statistics. As such, we should consider a similar presentation of the realistic results in Chap. 7, but of a qualitative nature. There are two different cases which need to be assessed, which are based upon the source of the empirical data that are used to compute the 2^{nd} order statistics. The first case is the one that was used as the basis of aggregated raw statistical data in Chap. 7. The second case is a truncation of the raw building

temperature time series dataset. The non-stationary night and weekend periods can be artificially extracted, along the same lines as the procedure used in Federspiel's re-calibration study [21, 22]. The advantage to doing this is that the temperature is regulated by a controller during the normal occupied periods. Therefore, control system error is the main contributor to building temperature variance during these periods, when the temperature is more "stationary." Different optimal P_b values will result from using the different methods of computing these statistics, and hence result in different behavior of the optimal alarm systems.

If we consider the realistic schedule according to Table 7.1, the optimal alarm system based on the "truncated" dataset is more appropriate for the "stationary periods" when the operating alarm systems are on-line, because that is the time period from which the data was extracted. The raw statistics used to design the alarm systems in Chap. 7 can *also* be used during the "occupied" stationary periods, but at an observed suboptimal disadvantage, because the aggregate statistics used for design were different than those during that period. However, using the statistics of the original dataset during the "non-stationary" period on HVAC start up, when the "arrival *and* operating" alarm systems are on-line will *not* result in as much of a suboptimal disadvantage. The reason for this is that the building temperature statistics during this "pre-occupied" period more accurately reflect the range of the modelled data.

Therefore, to summarize the main conclusions, let's first assume that the alarm systems have a positive utility for unautomated responses that consider both energy and discomfort rather than only complaint costs. Given this, design of the alarms systems corresponding to different time periods should be based on empirical data from the same time period. Specifically, for both the arrival and operating complaint period, optimal alarm design should be based on the raw empirical dataset used to design the alarm systems in Chap. 7. However, for the "operating only" complaint period, the truncated dataset should be used since it represents the counterpart "occupied" stationary period.

Future research areas pertaining to the use of automated or unautomated optimal alarm systems in general include several mentioned or alluded to thus far. They are summarized and augmented in the following list:

1. An empirical study needs to be conducted for this application that illustrates the clear superiority of using optimal alarm systems over the "simple" and "predictive" methods described in the introduction to Chap. 3. Svensson illustrates the superiority of optimal alarm systems to "predictive" alarm systems for single upcrossing events of a general nature [69, 70].

2. The economic value of discomfort (as in Mozer et.al. [52, 53]) needs further study and integration with optimal alarm systems, without requiring a tie to the complaint cost.

3. The sub-optimality involved in making stationary and Gaussian assumptions needs further investigation. This might include a possible study of non-stationary and non-Gaussian disturbances to develop more accurate predictions for the alarm system. Hence, instead of relying on fundamental spectral characteristics, true periodicity and trends in the data can be replicated and predicted.

4. An adaptive, real-time on-line implementation similar to the one used by Antunes et.al. [2] needs to be developed for optimal alarm systems as presented in this thesis. It should have the functionality of learning parameters during system execution. Another technique that might be considered is to use the E-M (Expectation-Maximization) algorithm, and exponential weighting for on-line system identification, similar to methods used by Weinstein et. al. [75]. The EM algorithm would need to appeal to mixed graphical modelling, where both continuous and discrete nodes would need to be used, as in Cowell et. al. [13] and Lerner et. al [41]. Because complaints can be modelled as binary discrete events, and there are other continuous processes involved, the hybrid nature of the problem reveals an additional layer of complexity.

5. Generalization of the theory in a mathematically rigorous fashion. The appeal of optimal alarm systems does not end with this application, and as such should be studied in depth.

Some examples are generalized by "Svensson's model," and all types of catastrophes that it might represent (environmental & economic systems such as seismic/earthquake applications, ranfall/runoff issues, & stock market crashes). Also, an example provided in Chap. 2 for this particular application was mentioned. Specifically, the fundamental idea of alarm systems can easily be extended to fault detection for specific HVAC system components and relevant monitored processes as proposed by Seem et. al. [64]. This also has obvious appeal when considering alarm-based control, using P_b as a tuning parameter. Recall that alarm-based control systems have a nice interpretation as fault detection systems for lower values of P_b, as discussed in Sec. 7.3. This idea may even have applicability for lighting satisfaction and balancing it with energy minimization. However, even more interestingly, the idea of using optimal alarm systems for fault detection can be studied in a more general context.

8.2 Optimal Control

Some of the chief disadvantages of implementing an optimal "alarm-based" control system to develop a thermostat setting policy are listed below.

- Sub-optimal results due to an approximated alarm region.
- Sensitivity of the alarm-based control response to the optimal value of P_b selected.
- Temporary discomfort to save energy during high electricity price periods, or for false alarms.
- Empirical data is not sufficient to implement alarm-based control. A model of the building temperature process which quantifies energy must be used. This model must be very accurate to allow for truly optimal predictions, alarms, and subsequent control actions to be based on reality.

However, all of the problems listed above have already been addressed with appropriate anecdotes or alternatives. Modelling considerations can be addressed with "future research area"

items #3 and #4 from the previous section. Primarily, the resolutions are based upon the design flexibility of the system, or alternate approximation methods. Recall from Sec. 7.2 that the design criteria, i.e. selecting the optimal P_b value, has several simultaneous interpretations. Tuning of the optimal alarm-based control system can proceed with several objectives in mind. Energy, maintenance and discomfort costs as well as sensitivity of the alarm system are all affected by tuning and selection of P_b. The advantages of being able to perform such analyses are unbounded, and clearly suitable candidates for further study. This is true not only for the present application, but for any applications where the utility of implementing new or using existing alarm systems and their integration with existing control systems has not been investigated.

For this particular building controls application, we now know that the management and automation of commercial building HVAC systems can benefit immensely from the insightful use of all available information sources. We've seen how modern HVAC systems using direct digital control provides useful performance data, and can interact synergistically with complaint statistics cultivated from modern maintenance management databases. The use of the latter to achieve useful practical objectives that are important to facility managers was virtually an untapped resource until recently. Application of the theory that drives some the of the newer techniques should present no barrier to the computational efficiency and practical usefulness of the resulting techniques. Therefore, the desired outcome, to make these new techniques more accessible to users of existing building control technology, and for facility managers to benefit from the results presented, can be achieved.

Bibliography

[1] I. Andresen and M.J. Brandemuehl. Heat storage in building thermal mass: A parametric study. *ASHRAE Transactions*, 98(1):910–918, Winter 1992.

[2] M. Antunes, A. Amaral Turkman, and K. F. Turkman. A Bayesian approach to event prediction. *Journal of Time Series Analysis*, 24(6):631–646, November 2003.

[3] ASHRAE. *ASHRAE/IESNA Standard 90.1-1989—Energy Efficient Design For New Buildings Except Low-Rise Residential Buildings.* American Society of Heating,Refrigerating,and Air Conditioning Engineers, Atlanta, GA, 1989.

[4] ASHRAE. *ASHRAE /ANSI Standard 55-1992, Thermal environmental conditions for human occupancy.* American Society of Heating,Refrigerating,and Air Conditioning Engineers, Atlanta, GA, 1992.

[5] ASHRAE. *1997 ASHRAE Handbook — Fundamentals.* American Society of Heating,Refrigerating,and Air Conditioning Engineers, Atlanta, GA, SI edition, 1997.

[6] Karl J. Åström. *Introduction to Stochastic Control Theory.* Academic Press, New York, 1970.

[7] John T. Barnett. *Zero-Crossing Rates of Some Non-Gaussian Processes with Application to Detection and Estimation.* PhD thesis, University of Maryland, October 1996.

[8] Stig-Inge Beckman, Jan Holst, and Georg Lindgren. Alarm characteristics for a flood warning

system with deterministic components. *Journal of Time Series Analysis*, 11(1):1–18, March 1987.

[9] Jack R. Benjamin and C. Allin Cornell. *Probability, statistics, and decision for civil engineers.* McGraw-Hill, New York, 1970.

[10] James E. Braun. Load control using building thermal mass. *Journal of Solar Energy Engineering*, 125(3):292–301, August 2003.

[11] J.E. Braun. Reducing energy costs and peak electrical demand through optimal control of building thermal storage. *ASHRAE Transactions*, 96(2):876–888, 1990.

[12] T. Y. Chen. Real-time predictive supervisory operation of building thermal systems with thermal mass. *Energy and Buildings*, 3(2):141–150, January 2001.

[13] Robert G. Cowell, Steffen L. Lauritzen, A. Philip David, and David J. Spiegelhalter. *Probabilistic Networks and Expert Systems.* Springer-Verlag, Inc., New York, 1^{st} edition, 1999.

[14] Harald Cramér and M.R. Leadbetter. *Stationary and Related Stochastic Processes.* John Wiley and Sons, 1967.

[15] Jacques DeMaré. Optimal prediction of catastophes with application to Gaussian processes. *Annals of Probability*, 8(4):840–850, August 1980.

[16] P. O. Fanger. *Thermal comfort: analysis and applications in environmental engineering.* McGraw-Hill, New York, 1972.

[17] Clifford C. Federspiel. *User-adaptable and minimum-power thermal comfort control.* PhD thesis, Massachusetts Institute of Technology, June 1992.

[18] Clifford C. Federspiel. Statistical analysis of unsolicited thermal sensation complaints in commercial buildings. *ASHRAE Transactions*, 104(1):921–923, 1998.

[19] Clifford C. Federspiel. Predicting the frequency of hot and cold complaints in buildings. *International Journal of HVAC&R Research*, 6(4):217–234, 2000.

[20] Clifford C. Federspiel. *Estimating the Frequency and Cost of Responding to Building Complaints*, chapter 56. Indoor Air Quality Handbook. McGraw-Hill, 2001.

[21] Clifford C. Federspiel, Rodney A. Martin, and Hannah Yan. Thermal comfort models and "call-out" (complaint) frequencies. Technical report, University of California, Berkeley, Center for the Built Environment, 2003.

[22] Clifford C. Federspiel, Rodney A. Martin, and Hannah Yan. Re-calibration of the complaint prediction model. *International Journal of HVAC&R Research*, 10(2), April 2004.

[23] G. E. Forsythe, M. A. Malcolm, and C. B. Moler. *Methods for Mathematical Computations*. Prentice-Hall, 1976.

[24] Marc Fountain, Gail Brager, Edward Arens, Fred Bauman, and Charles Benton. Comfort control for short-term occupancy. *Energy and Buildings*, 21(2):1–13, 1994.

[25] Gene F. Franklin, David Powell, and Michael L. Workman. *Digital Control of Dynamic Systems*. Addison-Wesley Co., 2nd edition, 1990.

[26] A.P. Gagge, A.P. Fobelets, and L.G. Berglund. A standard predictive index of the human response to the thermal environment. *ASHRAE Transactions*, 92(B):709–731, 1986.

[27] Alan Genz. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1:141–149, 1992.

[28] J. S. Haberl and S. Thamilseran. The great energy predictor shootout II: Measuring retrofit savings. *ASHRAE Journal*, 40(1):49–56, 1998.

[29] Maher Hamdi, Gérard Lachiver, and François Michaud. A new predictive thermal sensation index of human response. *Energy and Buildings*, 29:167–178, 1999.

[30] Anders Hansson. Control of level-crossings in stationary Gaussian random sequences. In *Proceedings of the 1992 American Control Conference*, pages 1065–1069, Chicago, Illinois, 1992.

[31] Anders Hansson. Control of level-crossings in stationary Gaussian random processes. *IEEE Transactions on Automatic Control*, 38(2):318–321, 1993.

[32] Anders Hansson. Optimal modifications of reference signals for critical processes using alarm signals. In *Preprints IFAC 12th World Congress*, pages 615–618, Sydney, Australia, 1993.

[33] Anders Hansson. Control of mean time between failures. *International Journal of Control*, 59(6):1485–1504, 1994.

[34] Gregor P. Henze, Robert H. Dodier, and Monced Krarti. Development of a predictive optimal controller for thermal energy storage systems. *International Journal of HVAC&R Research*, 3(3):233–264, July 1997.

[35] K. J. Hunt. *Polynomial Methods in Optimal Control and Filtering*. 49. Peter Peregrinus Ltd. on behalf of IEEE, 1993.

[36] ISO. *ISO 7730 Standard 1994 Moderate thermal environments—determination of the PMV and PPD indices and specification of the conditions for thermal comfort*. International Organization for Standardization, 1994.

[37] Michael I. Jordan. An introduction to probabilistic graphical models. Manuscript used for Class Notes of CS281A at UC Berkeley, Fall 2002.

[38] Benjamin Kedem. *Time Series Analysis by Higher Order Crossings*. IEEE Press, 1994.

[39] Bill Knox, Virginia Lew, Daryl Mills, and Michael Sloss. Energy accounting: A key tool in managing energy costs—energy efficiency project management handbook. Handbook, California Energy Commission, January 2000.

[40] K. F. Krieder and J. S. Haberl. Predicting hourly building energy usage: The great energy predictor shootout – overview and discussion of results. *ASHRAE Transactions*, 100(2):1104–1118, 1994.

[41] Uri Lerner, Eran Segal, and Daphne Koller. Exact inference in networks with discrete children of continuous parents. In *Proceedings of the 17th Annual Conference on Uncertainty in AI (UAI)*, pages 319–328, 2001.

[42] Geoffrey J. Levermore, David J. Lowe, and James W. Ure. Occupant feedback questionnaire producing a fingerprint and a score. *ASHRAE Transactions*, 105(2):661–670, 1999.

[43] Frank L. Lewis. *Optimal Control*. John Wiley & Sons, 1986.

[44] Craig Lin. Multiple wireless sensors with single HVAC system control. Master's thesis, University of California, Berkeley, 2002.

[45] Georg Lindgren. Optimal prediction of level crossings in Gaussian processes and sequences. *Annals of Probability*, 13(3):804–824, August 1985.

[46] Rodney A. Martin. Optimized response to thermal sensation complaints in buildings. Master's thesis, University of California, Berkeley, December 2000.

[47] Rodney A. Martin, Clifford C. Federspiel, and David M. Auslander. Responding to thermal sensation complaints in buildings. *ASHRAE Transactions*, 108(1):407–412, Winter 2002.

[48] Rodney A. Martin, Clifford C. Federspiel, and David M. Auslander. Supervisory control for energy savings and thermal comfort in commercial building HVAC systems. In *Proceedings of the AAAS 2002 Spring Symposium*, pages 67–74. American Association for Artificial Intelligence, AAAI Press, 2002.

[49] Mario J. Miranda and Paul L. Fackler. *Applied Computational Economics and Finance*. MIT Press, 2002.

[50] Jacques Miriel and Florence Fermanel. Classic wall gas boiler regulation and a new thermostat using fuzzy logic — improvements achieved with a fuzzy thermostat. *Applied Energy*, 68:229–247, 2001.

[51] Rob Moult. Fundamentals of DDC. *ASHRAE Journal*, 42(11):19–20,22–23, November 2000.

[52] Michael C. Mozer, Michael H. Coen, Richard Hasha, and James L. Flanagan. An intelligent environment must be adaptive. *IEEE Intelligent Systems and Their Applications*, 14(2):11–13, March-April 1999.

[53] Michael C. Mozer, M. I. Jordan, and T. Petsche. The neurothermostat: Predictive optimal control of residential heating systems. *Advances in Neural Information Processing Systems 9*, pages 953–959, 1997.

[54] Tatsuo Nagai. Optimization method for minimizing annual energy, peak demand, and annual energy cost through use of building thermal storage. *ASHRAE Transactions*, 108(1):43–53, 2002.

[55] Torbjörn Norlander and Pertti M. Mäkilä. A sampled-data theory for fixed structure LQ controller design. Research Report 07, Luleå Technical University, 2000.

[56] Athanasios Papoulis. *Probability, random variables, and stochastic processes.* McGraw-Hill series in electrical engineering. Communications and signal processing. WCB/McGraw-Hill, Boston, 3rd edition, 1991.

[57] Jim Pitman. *Probability.* Springer-Verlag, 1993.

[58] Andrew B. Price and Theodore F. Smith. Optimal control strategies for HVAC and building systems under the influence of building mass. In *Proceedings of the International Conference on Energy and Environment*, pages 194–204, Beijing, China, 1998. China Machine Press.

[59] A. J. Rainal. Zero-crossing principle for detecting narrow-band signals. *IEEE Transactions on Instrumentation and Measurement*, IM-15(1-2):38–43, 1966.

[60] T.A. Reddy, L.K. Norford, and W. Kempton. Shaving residential air-conditioner electricity peaks by intelligent use of the building thermal mass. *Energy*, 16(7):1001–1010, July 1991.

[61] S. O. Rice. Mathematical analysis of random noise. *Bell System Technology Journal*, 24:46–156, 1945.

[62] M.B. Russell and P.N. Surendran. Influence of active heat sinks on fabric thermal storage in building mass. *Applied Energy*, 70(1):17–33, September 2001.

[63] J.E. Seem and J.E. Braun. Adaptive methods for real-time forecasting of building electrical demand. *ASHRAE Transactions*, 97(1):710–721, Winter 1991.

[64] John E. Seem, John M. House, and Richard H. Monroe. On-line monitoring and fault detection. *ASHRAE Journal*, 41(7):21–26, July 1999.

[65] M.R. Shaw, K.W. Treadaway, and S.T.P. Willis. Effective use of building mass. *Renewable Energy*, 5(2):1028–1038, August 1994.

[66] Max Sherman. A simplified model of thermal comfort. *Energy and Buildings*, 8:37–50, 1985.

[67] Rahmat Shoureshi, Paul Torcellini, and Karim Rahmani. Intelligent occupant comfort control system. In *Winter Annual Meeting of the American Society of Mechanical Engineers*, volume 45, pages 105–111, New York, NY, USA., Winter 1992. ASME.

[68] Charles J. Stone. *A Course in Probability and Statistics*. Duxbury Press, 1996.

[69] Anders Svensson. *Event Prediction and Bootstrap in Time Series*. PhD thesis, Lund Institute of Technology, September 1998.

[70] Anders Svensson, Jan Holst, R. Lindquist, and Georg Lindgren. Optimal prediction of catastrophes in autoregressive moving-average processes. *Journal of Time Series Analysis*, 17(5):511–531, 1996.

[71] Jason Teeter and Mo-Yuen Chow. Application of functional link neural network to HVAC thermal dynamic system identification. *IEEE Transactions on Industrial Electronics*, 45(1):170–176, February 1998.

[72] A.J.N. van Breemen and T.J.A. de Vries. Design and implementation of a room thermostat using an agent-based approach. *Control Engineering Practice*, 9:233–248, 2001.

[73] Harry L. VanTrees. *Detection, estimation, and modulation theory.* J. Wiley, 1992.

[74] J.C. Vischer. Using occupants' experiences to monitor indoor air quality in office buildings. *ASHRAE Transactions*, 99(2):1111–1115, 1993.

[75] Ehud Weinstein, Alan V. Oppenheim, Meir Feder, and John R. Buck. Iterative and sequential algorithms for multisensor signal enhancement. *IEEE Transactions on Signal Processing*, 42(4):846–859, April 1994.

[76] C.P. Yaglou and P. Drinker. The summer comfort zone: Climate and clothing. *Journal of Industrial Hygiene and Toxicology*, 10:350–363, 1928.

[77] Radu Zmeureanu and Paul Fazio. Role of computers in the design of energy-responsive buildings. *Canadian Journal of Civil Engineering*, 16(1):55–61, February 1989.